

# The Effect of Socio-economic Status on Outcomes in Cystic Fibrosis

Thesis submitted in accordance with the requirements of the  
University of Liverpool for the degree of Doctor in Philosophy by

David Carlton Taylor-Robinson

June 2013

Department of Public Health and Policy  
University of Liverpool

## **Declaration**

This thesis is my own work. The material contained in this thesis has not been presented, nor is currently being presented, either wholly or in part for any other degree or qualification.

## **Dedication**

I would like to dedicate this thesis to my dad, Dr Carlton Hugh Taylor-Robinson. His passion for science has had a lasting influence on me. He was a meticulous medical researcher, whose own doctorate was never completed owing to his unfailing dedication to his sons. He believed wholeheartedly in social equality and the NHS and this would have been a topic of great interest to him. Without his encouragement I would never have learned to throw myself into academic work. I wish he could read this.

## **Acknowledgements**

The studies in this thesis were funded by the Medical Research Council (MRC) through a Population Health Scientist Fellowship.

There are many people who have helped me during my Fellowship and doctoral studies:

I am particularly grateful to my fantastic supervisors/friends, from whom I have learnt so much. To Peter Diggle, for his patience, and for always having the answer. To Margaret Whitehead, who always conveys her firm belief in social justice with such grace – a constant inspiration. To Ros Smyth for keeping things simple and focussed on children with CF. I shall miss our stimulating meetings.

I'd like to thank everyone at the CF Trust, especially Elaine Gunn and Matt Baker, for their help with the registry, and the clinicians in the UK and Denmark who have carefully collected the data.

I'd like to thank the people behind the numbers – all the children and adults with CF whose lives are captured in this thesis.

Thanks to friends and colleagues in Denmark: Tania Pressler, Finn Diderichsen, Karsten Thielen, Hanne Verbert Olesen, and Oluf Schiøtz.

Thanks to my family for all their love and support: To Sayling and Marie, John and Katie, and Matt and Cate for helping out at various times. Thanks especially to Viktoria, without whom this would have been impossible.

Lastly, thanks to Kate, Mia, Tom, and Carl, for always being there.

## Abstract

**Introduction:** Reducing inequalities in health is a public health imperative. In the UK and internationally policies are being implemented to try to reduce health inequalities, with limited success. This thesis examines the effect of socio-economic status (SES) on clinical outcomes, healthcare use and employment opportunities in people with cystic fibrosis (CF). Poorer socio-economic circumstances have been linked with worse outcomes in cystic fibrosis. Because CF is genetically determined, this offers an opportunity to investigate the impact of SES on health and social outcomes, in a chronic condition *without a socio-economic gradient in incidence*. This provides a useful case for understanding how health inequalities are generated, in order to develop more effective interventions, for people with CF and more generally.

**Methods:** I analyse, for the first time, the national CF registers from the UK and Denmark, using longitudinal modeling techniques. Mixed-effects models are used to assess the association between measures of SES and longitudinal outcomes, adjusting for clinically important covariates. Study 1 explores longitudinal weight, height, BMI, %FEV<sub>1</sub>, risk of *Pseudomonas* colonisation, and the use of major CF treatment modalities, and their association with small-area deprivation (8055 people with 49,337 observations between 1996 and 2010). Study 2 explores longitudinal employment status in adults with CF in the UK, and its association with small-area deprivation, disease severity, and time in hospital. Study 3 presents a novel longitudinal analysis of the Danish CF registry (70,448 %FEV<sub>1</sub> measures on 479 patients seen monthly between 1969 and 2010), to understand the way %FEV<sub>1</sub> changes over time.

**Results:** Compared with the least deprived areas in the UK, children with CF from the most deprived areas weighed less, were shorter, had a lower body-mass index, were more likely to have chronic *P. aeruginosa* infection, and have a lower %FEV<sub>1</sub>. These inequalities were apparent very early in life and did not widen thereafter. On a population level, after adjustment for disease severity, children in the most deprived quintile were more likely to receive intravenous antibiotics and nutritional treatments compared with individuals in the least deprived quintile. Patients from the most disadvantaged areas were less likely to receive DNase or inhaled antibiotic treatment. In adults deprivation, disease severity, and time in hospital all influence employment chances in CF. Furthermore, deprivation amplifies the harmful effects of disease severity on employment: the employment chances of people with CF with poor lung function from disadvantaged areas are damaged to a greater extent than for their counterparts living in the least disadvantaged circumstances. The Danish analysis quantifies the short-term variability in %FEV<sub>1</sub> (SD 6.3%) and shows that lung function measures are correlated for over 15 years.

**Conclusions:** In the UK, children with CF from more disadvantaged areas have worse growth and lung function compared with children from more affluent areas, but these inequalities do not widen with advancing age. Policies to reduce inequalities should thus focus on the early years. Clinicians consider deprivation status, as well as disease status, when making decisions about treatments, and this might mitigate some effects of social disadvantage. The *differential social consequence* of having CF in terms of employment is likely to be an important pathway for the amplification of health inequalities.

# Contents

Figures .....	viii
Tables .....	xi
Abbreviations .....	xiii
Chapter 1: Introduction .....	1
Relevance of the issue .....	1
Previous research and gaps in the literature .....	2
Objectives of the research .....	5
Structure of this thesis .....	6
Chapter 2: Literature review .....	9
The policy context to health inequalities .....	11
The effects of childhood disadvantage .....	14
Understanding the causes of health inequalities .....	19
Studying tracer groups to better understand pathways to health inequalities .....	33
CF as a tracer condition .....	34
Key features of CF .....	38
Health inequalities and CF .....	56
Gaps in the literature .....	61
Summary .....	62
Chapter 3: Methods .....	64
Source and extraction of UK SES data .....	70
Danish dataset .....	74
Contrasting UK and Danish datasets .....	78
Statistical methods .....	80
Pilot study and consideration of sample size .....	101
Computation .....	103
Chapter 4: Study 1 – The effect of social deprivation on clinical outcomes and the use of treatments in the UK CF population: A longitudinal study .....	104
Abstract .....	104
Introduction .....	105
Methods .....	106
Results .....	108
Discussion .....	151
Interpretation .....	156
Chapter 5: Study 2 – A longitudinal study of the impact of social deprivation and disease severity on employment status in the UK CF population .....	158
Abstract .....	158
Introduction .....	160
Methods .....	162
Results .....	166
Discussion .....	178
Interpretation .....	181
Chapter 6: Study 3 – Understanding the natural progression in %FEV <sub>1</sub> decline in patients with CF: A longitudinal study .....	182
Abstract .....	182
Introduction .....	183
Methods .....	184
Results .....	187
Interpretation .....	208

Chapter 7: Discussion .....	209
Introduction .....	209
Key findings with reference to objectives.....	209
How has this thesis contributed to the literature? .....	216
Contribution to our understanding of CF epidemiology .....	216
Contribution to the knowledge base around health inequalities .....	246
Methodological advances.....	253
Critique of overall study design .....	255
What are the implications for policy and clinicians? .....	263
Conclusions .....	268
Further research.....	268
Research currently in progress from this thesis .....	269
References .....	273
Appendix 3 (Pertaining to Chapter 3, Methods) .....	305
Appendix 4 (Pertaining to Chapter 4) .....	325
Weight .....	325
Height.....	328
BMI .....	332
%FEV <sub>1</sub> .....	336
<i>Pseudomonas</i> colonisation.....	338
Any IV antibiotic therapy.....	340
Any home IV therapy.....	341
Nutritional therapy .....	343
DNase.....	344
Any inhaled therapy <18.....	345
Illustrative regression diagnostics .....	346
Appendix 5 (Pertaining to Chapter 5) .....	348
Appendix 6 (Pertaining to Chapter 6) .....	349
Appendix (Publications from this thesis).....	354

## Figures

Figure 1: Schematic overview of the studies in this thesis .....	8
Figure 2: Low birth weight; asthma admission rate; GCSE success rate; and life expectancy versus child poverty for wards in Liverpool 2011 .....	15
Figure 3: Children achieving a good level of development at age 5 versus child poverty.....	18
Figure 4: The main influences on health.....	21
Figure 5: The social determinants of health inequalities.....	22
Figure 6: A conceptual model for studying the effect of social position and social context on health .....	31
Figure 7: Median FEV <sub>1</sub> % predicted by paediatric centre/clinic all centres and networks. ....	46
Figure 8: Flowchart of included participants for weight analysis.....	68
Figure 9: Data follow-up in the population aged <40 years.....	69
Figure 10: Year of birth for people included in the weight analysis 1996 to <2010 ..	69
Figure 11: Density plot comparing distributions of deprivation scores .....	73
Figure 12: Cumulative density plot comparing deprivation score distribution for the whole UK population versus the CF population .....	74
Figure 13: Map of Denmark showing location of CF centres.....	75
Figure 14: Histogram of year of birth for Danish dataset .....	77
Figure 15: Histogram of frequency and duration of follow up .....	78
Figure 16: Twenty randomly selected %FEV <sub>1</sub> profiles from the UK CF registry for people aged<20 .....	79
Figure 17: Five randomly selected %FEV <sub>1</sub> profiles from the Danish registry .....	79
Figure 18: PubMed results containing “longitudinal” by year, 1960-2012 .....	80
Figure 19: Spaghetti plot for %FEV <sub>1</sub> versus age illustrating the longitudinal nature of the data. ....	88
Figure 20: Random intercept and slope model for %FEV <sub>1</sub> decline .....	91
Figure 21: A typical example of a theoretical variogram.....	95
Figure 22: Simulated data illustrating the unbiased estimates resulting from a mixed model analysis, where the data is assumed to be MAR, compared to the GEE estimate.....	100
Figure 23: Scatterplot of % predicted FEV <sub>1</sub> versus age with smoothed mean .....	102
Figure 24: Logic model for %FEV <sub>1</sub> analysis .....	107
Figure 25: Logic model for IV therapy analysis .....	108
Figure 26: Distribution of incident cases by deprivation quintile.....	109
Figure 27: Kaplan-Meier plot of time to diagnosis by deprivation quintile.....	111
Figure 28: Anthropometric outcomes: mean cross-sectional weight, height, and BMI by age comparing extremes of deprivation quintile (red most deprived). .....	114
Figure 29: Modelled growth trajectories for children, comparing least (blue) and most deprived quintiles (red).....	115
Figure 30: Spaghetti plot of weight SD score versus age in paediatric age group... ..	116
Figure 31: Piecewise modelling approach to weight z score trajectory .....	117
Figure 32: Exploratory analysis showing smoothed means of weight for age SD score versus age, stratified by covariates, for people<18 .....	118
Figure 33: Deprivation quintile contrast from final weight model <18 .....	119
Figure 34: Weight for age Z score versus age, illustrating the effect of sex, screening status, delta F508 carrier status, and ethnicity .....	121

Figure 35: Piecewise modelling approach to weight SD score trajectory in analysis stratified by screening status .....	122
Figure 36: Deprivation effect on weight in screening stratified analysis.....	123
Figure 37: Spaghetti plot of weight for age SD score versus age in adult age group .....	124
Figure 38: Exploratory analysis showing smoothed means of weight for age SD score versus age, stratified by covariates, for people>18 .....	125
Figure 39: Deprivation quintile contrast from final weight model >18.....	125
Figure 40: Weight for age versus age, illustrating covariate contrasts from the final longitudinal models.....	127
Figure 41: Height for age versus age, illustrating covariate contrasts from the final longitudinal models.....	128
Figure 42: BMI for age versus age, illustrating covariate contrasts from the final longitudinal models.....	129
Figure 43: BMI for age versus age, illustrating covariate contrasts from the final longitudinal models.....	130
Figure 44: Respiratory outcomes: Mean cross-sectional %FEV <sub>1</sub> and <i>P. aeruginosa</i> colonization prevalence by age comparing extremes of deprivation quintile..	131
Figure 45: %FEV <sub>1</sub> versus age from 5 to 40 .....	132
Figure 46: GAMs showing the shape of the relationship between %FEV <sub>1</sub> and deprivation score .....	133
Figure 47: %FEV <sub>1</sub> trajectories, illustrating the effect of deprivation, sex, screening, genotype, ethnicity, <i>Pseudomonas</i> , CFRD and BMI status .....	135
Figure 48: %FEV <sub>1</sub> versus age in the >18 group, illustrating sex and genotype effect .....	136
Figure 49: Cross sectional and modelled longitudinal <i>Pseudomonas</i> prevalence by deprivation.....	137
Figure 50: GAMs showing the shape of the relationship between risk of <i>Pseudomonas</i> colonisation and deprivation score.....	138
Figure 51: Use of therapies: any IV antibiotic therapy, home IV antibiotic therapy, hospital IV antibiotic therapy, supplemental feeding, DNase, and inhaled antibiotics, by age comparing extremes of deprivation quintile .....	140
Figure 52: Cross sectional and modelled longitudinal percentage of people using any IV therapy in the preceding year by deprivation in the <18 and >18 age group .....	142
Figure 53: GAMs showing the shape of the relationship between log-odds of any IV therapy and deprivation score. ....	142
Figure 54: Scatterplot of log IV days (home and hospital) versus age, with fitted longitudinal trajectory by deprivation quintile.....	144
Figure 55: Cross sectional and modelled longitudinal percentage of people using any nutritional therapy in the preceding year by deprivation in the <18 age group and <18 age group.....	145
Figure 56: Cross sectional percentage of people using any DNase therapy, and inhaled antibiotic, by deprivation quintile .....	146
Figure 57: Cross-sectional full time and part time employment, by deprivation and sex .....	163
Figure 58 Logic model to inform analysis of employment status.....	164
Figure 59: Cross sectional employment prevalence by age and year .....	169
Figure 60: GAMs for risk of employment .....	170
Figure 61: Empirical logit plots by age and SES for females in the dataset.....	171

Figure 62: Longitudinal employment trajectory versus age of people with CF in UK CF Registry, by deprivation quintile, sex, %FEV <sub>1</sub> , BMI SD score and days in hospital.....	174
Figure 63: Exploratory analysis of interaction effect between deprivation and %FEV <sub>1</sub> on employment status.....	175
Figure 64: Longitudinal employment trajectory versus age, demonstrating the interaction between deprivation and %FEV <sub>1</sub> .....	177
Figure 65: Scatterplot of %FEV <sub>1</sub> versus age in Denmark, with a randomly selected individual highlighted in each panel.....	188
Figure 66: All Danish data with smoothed mean trend.....	189
Figure 67: Residual plot for individual in Danish population.....	190
Figure 68: Quantifying the variability in %FEV <sub>1</sub> with the variogram approach.....	191
Figure 69: Proportion of variability in an individual's %FEV <sub>1</sub> at follow-up time t that is explained by their %FEV <sub>1</sub> at baseline.....	192
Figure 70: Comparison between empirical variogram and MLE variogram estimate.....	193
Figure 71: Scatterplot of standardised residuals versus fitted values.....	194
Figure 72: Simulated realisations from the final model.....	194
Figure 73: Comparison of conventional random-intercept and slope model over short and long follow-up periods, versus proposed Gaussian process model.....	197
Figure 74: Univariate effect of pancreatic status.....	200
Figure 75: Effect of key covariates on %FEV <sub>1</sub> .....	201
Figure 76: Scatterplot of data for post-1998 cohort.....	202
Figure 77: Scatterplot of data for 1988-1998 cohort.....	202
Figure 78: Logic model based on the Diderichsen model outlining exposures for poor outcomes in CF.....	228
Figure 79: Effect of sex on lung function decline in the US and the UK.....	237
Figure 80: How risk reduction and health promotion strategies influence health development.....	249
Figure 81: Effect of parental education level on lung function decline in Denmark.....	270
Figure 82: Joint modelling of lung function and survival outcomes.....	271
Figure 83: Spaghetti plot of height for age SD score versus age in paediatric age group.....	328
Figure 84: Exploratory analysis showing smoothed means of height for age Z score versus age, stratified by covariates.....	328
Figure 85: Spaghetti plot of height for age Z score versus age in adult age group.....	329
Figure 86: Exploratory analysis showing smoothed means of height for age Z score versus age, in adult age group, stratified by covariates.....	329
Figure 87: Spaghetti plot of BMI for age Z score versus age in paediatric age group.....	332
Figure 88: Exploratory analysis showing smoothed means of BMI for age Z score versus age, stratified by covariates.....	332
Figure 89: Spaghetti plot of BMI for age Z score versus age in adult age group.....	333
Figure 90: Exploratory analysis showing smoothed means of BMI for age Z score versus age, in adult age group, stratified by covariates.....	333
Figure 91: Exploratory analysis showing smoothed means of %FEV <sub>1</sub> score versus age, stratified by covariates.....	336
Figure 92: %FEV <sub>1</sub> < 18 years, final model.....	346
Figure 93: Any IV therapy < 18 years, final model.....	347

## Tables

Table 1: National CF registries .....	50
Table 2: Domains and weights used to generate IMD scores for UK constituent countries .....	72
Table 3: Unadjusted characteristics of study population by deprivation quintile: UK CF registry 1996-2010 .....	110
Table 4: Cox regression of time to diagnosis by deprivation quintile .....	111
Table 5: Summary of adjusted effects of deprivation on clinical outcomes and use of therapies in the paediatric and adult CF population in the UK .....	112
Table 6: Final linear mixed-effects regression models for growth in the <18 age group .....	120
Table 7: Final linear mixed-effects regression models for growth in the >18 age group .....	126
Table 8: Final regression models for %FEV <sub>1</sub> in <18 age group .....	134
Table 9: Final regression models for any IV therapy in 5 to <18 age group .....	141
Table 10: Comparison of characteristics of eligible population versus those not meeting the inclusion criteria .....	147
Table 11: Additional models fitted as robustness tests, based upon the final FEV <sub>1</sub> model .....	149
Table 12: Regression coefficients from explanatory models fitting deprivation as a five-level factor .....	150
Table 13: Characteristics of study population in UK CF Registry by employment status at baseline .....	166
Table 14: Baseline characteristics at entry to the dataset, stratified by deprivation quintile .....	167
Table 15: Log-odds for the final nested GLMMs .....	173
Table 16: Percentage of people with CF in employment at age 30 .....	176
Table 17: Baseline characteristics of Danish CF population .....	187
Table 18: Univariate associations between covariates and %FEV <sub>1</sub> .....	199
Table 19: Estimates from final multivariate model .....	201
Table 20: Final regression models for weight SD score in <18 age group .....	325
Table 21: Final regression models for weight SD score in >18 age group .....	327
Table 22: Final regression models for height SD score in <18 age group .....	330
Table 23: Final regression models for height SD score in >18 age group .....	331
Table 24: Final regression models for BMI SD score in <18 age group .....	334
Table 25: Final regression models for BMI SD score in >18 age group .....	335
Table 26: Final regression models for %FEV <sub>1</sub> in >18 age group .....	337
Table 27: Final regression models for <i>P. aeruginosa</i> colonisation in <18 age group .....	338
Table 28: Final regression models for <i>P. aeruginosa</i> colonisation in >18 age group .....	339
Table 29: Final regression models for any IV therapy age 18 to <40 .....	340
Table 30: Robustness test: regression models for use of any IV therapy 5 to <18 ..	341
Table 31: Final regression models for any home IV therapy age 5-18 .....	341
Table 32: Final regression models for any home IV therapy age 18-40 .....	342
Table 33: Final regression models for amount of any IV therapy age 5 to <18 .....	342
Table 34: Final regression models for amount of IV therapy age 18 to <40 .....	343
Table 35: Final regression models for any nutritional support <18 age .....	343

Table 36. Final regression models for any nutritional support age 18-40.....	344
Table 37. Final regression models for any DNase <18 age .....	344
Table 38: Final regression models for any DNase age 18-40 .....	345
Table 39: Final regression models for any inhaled therapy <18 age .....	345
Table 40: Final regression models for any inhaled therapy 18-40 years .....	346
Table 41: Log odds for the final GLMMs, with added educational variable.....	348

## Abbreviations

%FEV <sub>1</sub>	Percent predicted forced expiratory volume in one second
ADHD	Attention deficit hyperactivity disorder
ANOVA	Analysis of variance
BAL	Bronchoalveolar lavage
BMI	Body mass index
BMI SD	Body mass index standard deviation
CCG	Clinical commissioning group
CF	Cystic fibrosis
CFQ-R	Cystic fibrosis questionnaire-revised
CFRD	Cystic fibrosis related diabetes
CFTR	Cystic fibrosis transmembrane conductance regulator
CI	Confidence Interval
COPD	Chronic obstructive pulmonary disease
CVD	Cardiovascular disease
DNase	Deoxyribonuclease
ESCF	Epidemiological study of cystic fibrosis
ETS	Environmental tobacco smoke
FEV <sub>1</sub>	Forced expiratory volume in one second
GAM	Generalized additive model
GDP	Gross domestic product
GCSE	General certificate of secondary education
GEE	Generalized estimating equations
GLM	Generalized linear model
GLMM	Generalized linear mixed model
HAZ	Height for age Z score or SD score
HR	Hazard ratio
HRQOL	Health related quality of life
ICD	International classification of disease
IMD	Index of multiple deprivation
IMR	Infant mortality rate
ISCED	International standard classification of education
IV	Intravenous
LDA	Longitudinal data analysis
LLSI	Limiting long-standing illness
LLSOA	Lower layer super output area
LME	Linear mixed effects
MANOVA	Multivariate analysis of variance
MAR	Missing at random
MCAR	Missing completely at random
MCS	Millenium cohort study
MDT	Multidisciplinary team
ML	Maximum likelihood
MRC	Medical research council
NHS	National health service
NMAR	Not missing at random
NS-SEC	National statistics socio-economic classification
OECD	Organisation for Economic Co-operation and Development
OLS	Ordinary least squares
ONS	Office of national statistics
OR	Odds ratio
PA	Pseudomonas aeruginosa
PI	Pancreatic insufficiency
POP	Pulmonary outcome prediction score
RCT	Randomized control trial
RIS	Random intercept and slope
SD	Standard deviation
SDH	Social determinants of health
SEP	Socio-economic position

SES	Socio-economic status
SGP	Stationary Gaussian process
TGF	Transforming growth factor $\beta 1$ (CF modifier gene)
TOBI®	Tobramycin inhalation solution
TPN	Total parenteral nutrition
UKCFS	UK cystic fibrosis survey
WAZ	Weight for age Z-score or weight SD score
WHO	World Health Organization

# Chapter 1: Introduction

## Relevance of the issue

Studies across the world have consistently shown that people from socio-economically disadvantaged backgrounds experience worse health than those in more socio-economically advantaged positions. These inequalities in health outcomes are preventable, amenable to policy intervention, and they are unjust. In the UK and internationally, policies are being implemented to try to reduce health inequalities, with limited success, since for many important health outcomes, such as life-expectancy in the UK, inequalities appear to be widening. In order to develop more effective interventions we need a better understanding of how, and when, these health differences are generated and maintained.

There is evidence of differences in survival for people with cystic fibrosis (CF) by socio-economic status (SES). Understanding the mechanisms by which these differences in survival are generated is important to improve care in CF. Furthermore, studying people with CF offers a unique opportunity to develop the evidence base around pathways to health inequalities more generally. Because CF is genetically determined, a socio-economic difference in incidence is not expected, but there may be important differences in outcomes over the course of people's lives. Studying the health of people with CF thus provides a valuable case study to investigate the impact of SES on health and social outcomes, in a chronic condition *without a socio-economic gradient in incidence*. Despite this 'level playing field' at the outset, the processes by which differential outcomes in CF are generated are unclear. Elucidating these mechanisms is important to inform policies, both to improve care in CF, and also to reduce inequalities in health more broadly.

## **Previous research and gaps in the literature**

There is a developing literature on the relationship between deprivation and poor outcomes for people with CF. As the commonest serious inherited disease in white populations, CF usually leads to premature death as a result of progressive lung infection and subsequent respiratory failure. Nowadays, we usually diagnose children with CF within in the first year of life, and they require intensive support from family and health care services subsequently. The improvement in survival over successive birth cohorts has been striking. Up until the 1950s, the majority of children with CF did not survive beyond infancy, but survival has increased dramatically since then, such that in the UK, estimates suggest children born in the 21<sup>st</sup> century have a median survival of over 50 years of age (Dodge et al., 2007). These improvements in survival, however, have not been shared equally, and early studies in the UK and the US have indicated patterning of survival by SES.

People with CF from more disadvantaged groups in the UK and US die earlier than those from more advantaged backgrounds. In the UK, the first study to look at social differences in survival in CF found a consistent trend from 1959 to 1986 towards higher age at death in CF patients in more advantaged social groups on the basis of occupational class (Britton, 1989). The authors speculated that the observed lower survival chances in lower social classes could relate to lack of access to the appropriate services, poorer nutrition, increased parental smoking, and poorer quality housing. Evidence from the US corroborated these findings, and indicated higher survival rates in the 1980s and 1990s among more advantaged socio-economic groups, measured by Medicaid status and area-based income, compared with their less advantaged counterparts (Schechter et al., 2001, O'Connor et al., 2003, Schechter, 2004). For instance the adjusted risk of death was around four times higher in CF patients with Medicaid cover, a surrogate for poverty, compared to those without Medicaid cover (Schechter et al., 2001). Having demonstrated a relationship with survival, subsequent studies predominantly from the US, have focussed on the potential mediating effect of health services on outcomes.

The role of health services in the generation of inequalities in CF remains unclear. In many chronic illnesses, differences in access to specialist health care by SES are evident. These differences often contribute to the exacerbation of health inequalities,

since a bias towards more advantaged groups is widespread across health care (Stirbu et al., 2011). This could be particularly important in the context of inequalities in outcomes for a disease like CF, where treatment advances have had such a profound effect on survival. The studies exploring access to services and treatments from the US, however, have demonstrated a mixed picture, depending on the type of treatment under consideration, and the exact measure of SES used (Schechter et al., 2009, Schechter et al., 2011). For instance, one study showed greater access to intravenous (IV) antibiotics for older children, on the basis of area-based income, but also showed no difference in access to DNase, another important treatment in CF, using Medicaid status as a measure of SES (Schechter et al., 2009). Overall the US studies have played down the role of health service factors in the generation of health disparities in CF. However, it is debatable how much of this evidence can be generalized to other health care settings, especially to the UK context, where access to health services is free at the point of use.

The literature on differential outcomes in CF in the UK is still in its infancy. To date, the research on the effect of SES on outcomes in CF in the UK is limited to studies of survival, using routinely collected death registration data (Britton, 1989, Barr HL, 2011). It remains unclear how these differences in survival in CF in the UK are generated, and at what stage in people's lives inequalities in health outcomes start to develop. For instance, in the early years, we can characterize CF by poor growth and under-nutrition, and a great deal of emphasis is placed on early diagnosis, and subsequent close growth monitoring and dietary supplementation. However, there is no data regarding socio-economic differences in early growth in the UK, though cross-sectional analyses from the US have suggested important differences in growth outcomes (Schechter et al., 2001). One can say the same for the acquisition of important lung infections, and lung function, which is perhaps the most important clinical outcome in CF, since it has been linked to survival. Thus there remain important gaps in our knowledge about the effect of SES on key clinical outcomes, which researchers have not studied in a systematic way over the life-course, in both paediatric and adult populations. Furthermore, the role of the National Health Service (NHS) in the UK in mitigating or potentiating any inequalities in outcomes remains unclear.

No-one has studied the effect of SES on social outcomes in CF, such as employment status. Understanding the *social consequences* of ill health is a key step in elucidating the pathways to health inequalities, since any adverse social outcomes that occur as a result of ill health can feedback and further damage health status. For example, ill health might lead to job loss, which can then have further adverse effects on health status, mediated by a range of physical and psycho-social mechanisms. Furthermore, there is evidence from social epidemiological studies to suggest that, in many settings, it is people from more disadvantaged backgrounds who particularly experience adverse health outcomes as a result of ill health – so called *differential social consequences*. What affect this pathway has on the generation and maintenance of health inequalities in CF remains an unstudied question. This gap in the literature is of particular concern, given that most people with CF born today will survive well into adulthood, and employment status is becoming an increasingly important consideration in the CF literature.

The studies in this thesis aim to fill these important gaps in our understanding of the social patterning of outcomes in CF, by utilizing rich longitudinal datasets from the UK and Denmark, to investigate differences in a range of longitudinal clinical, and social outcomes by SES. Furthermore, a key question in this thesis is to clarify the importance of health care and the role of the NHS, in the pathways to differential outcomes in people with CF. Importantly, the findings of these studies are intended to produce generalizable insights into the influence of social factors on chronic disease outcomes over the life-course, and the influence of child health on later adult health. The final step is to develop recommendations for addressing any inadequacies in CF care or public policy that are uncovered.

## **Objectives of the research**

Following a review of the literature, this thesis proceeds from the assumption that low SES has an adverse effect on survival for people with CF. The overall aim of this thesis is to investigate the effect of SES on clinical outcomes, healthcare use and employment opportunities in people with CF. To this end the main objectives of this thesis are to answer the following questions:

In people with CF:

1. What is the relationship between SES and important longitudinal clinical outcomes, such as growth, and lung function?
2. What is the relationship between SES and health service use?
3. What is the impact of SES on longitudinal employment status?
4. What are the policy implications of the above?

At the outset of this thesis my intention was to analyse data from the Danish CF register, using similar approaches to those used to analyse the UK data, but utilising the individual level SES data available in Denmark through data linkage. However, due to the very frequent monthly follow-up of CF patients in Denmark, it was necessary to develop a new approach to analysing the Danish dataset, and thus the focus of the Danish analysis presented in this thesis is around the methodological challenges of modelling lung function decline in the Danish population. The objective of the Danish study is to describe a novel method to model how %FEV<sub>1</sub> changes over time, in a way that is useful for clinicians at both the individual and the population level. Using this approach, research on the effect of SES on CF outcomes in Denmark is on-going.

## Structure of this thesis

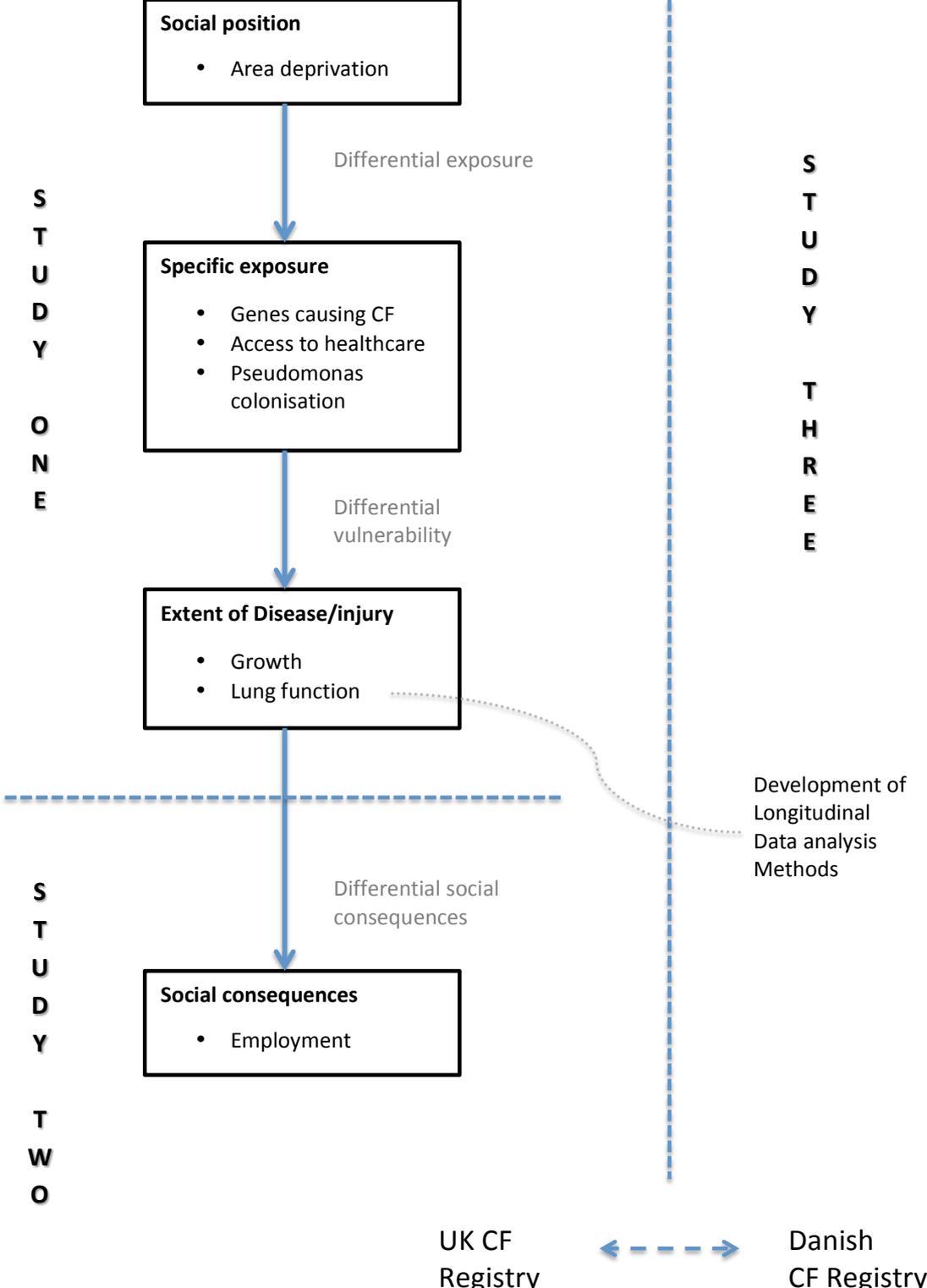
Figure 1 provides a schematic overview of the three studies in this thesis.

- Chapter 2 (Literature review) provides an introduction to health inequalities, and how these are generated. Then relevant aspects of CF are reviewed, before moving on to health inequalities in CF. The chapter finally highlights key gaps in knowledge.
- Chapter 3 (Methods) describes the UK and Danish CF datasets and the longitudinal data analysis techniques used in the thesis
- Chapter 4 (Results) presents and discusses the results of study 1. This is the longest of the results chapters, and focusses on an analysis of the effect of SES on clinical outcomes and health care use in the UK CF population. This addresses objectives 1 and 2 above, and also generates an in-depth description of longitudinal outcomes in CF in the UK.
- Chapter 5 (Results) presents and discusses the results of study 2, in which I focus on employment outcomes in the UK CF population, and look for differential social consequences by SES.
- Chapter 6 (Results) presents and discusses the results of study 3, in which I explore the dynamics of lung function decline in people with CF in Denmark, and apply a novel modelling approach.
- Chapter 7 (Discussion) pulls together the key findings and themes from Studies 1 to 3. I then describe the contribution of the studies to the existing literature, which extends beyond the initial objectives, in three main areas: CF epidemiology; our understanding of health inequalities more generally; and specific methodological advances. There is then a discussion of the strengths and limitations of the overall approach and study design. I address Objective 4 of the thesis with a section on the policy implications arising

from the studies, and the chapter concludes with a description of on-going research.

- There are additional appendices containing the data entry screens for the UK registry additional tables and plots to go with the analysis chapters, some R programming code, and finally the main publications from this thesis. Appendix 3 relates to Chapter 3, and so forth. There is no appendix 1 or 2.

**Figure 1: Schematic overview of the studies in this thesis**



## Chapter 2: Literature review

In Liverpool in the 1830s, a general practitioner named William Henry Duncan undertook a survey into the living conditions of parents in the city centre area. The population of Liverpool had boomed to almost one million, from 60,000 in 1770, putting huge pressure on housing in the city, and he had observed a concurrent rise in the prevalence of infectious diseases in the population, including tuberculosis (TB), whooping cough, and cholera. His survey showed that about one third of the population was living in the cellars of small back to back houses, with up to 16 people occupying each room (Ashton and Seymour, 1988). Describing the situation of a woman who had recently given birth to her child, Duncan wrote (Wohl, 1983):

*“I visited a poor woman in distress...she had been confined only a few days, and herself and infant were lying on straw in a vault...with a clay floor impervious to water. There was no light or ventilation and the air was dreadful. I had to walk on bricks across the floor to reach her bedside, as the floor itself was flooded with stagnant water...There are hordes of poor creatures living in cellars which are almost as bad and offensive as charnel-houses”*

Duncan went on to be appointed Liverpool’s medical officer for health, the first public health appointment of its kind in the world. His further studies, before the widespread use of statistical techniques such as standardization, documented differences in the in average age at death in 1848, comparing Liverpool, with an average of 19 years, to rural Wiltshire, at 36.5 years (Ashton, 1988). This was at a time when the first use of death registration statistics also showed shocking differences in the average age at death for different sections of the population. For instance, average age at death for ‘Gentry and professional’ classes in Rutland was 52 years compared to 15 years for ‘labourers and artisans’ in Liverpool (Lancet, 1843). Duncan’s work clearly linking adverse living conditions with poor health outcomes, along with that of other social reformers such as Edwin Chadwick, informed a public health movement, culminating in widespread sanitary reforms, and

government legislation such as the National Public Health Act in 1848 (Frazer, 1947).

Differences in life expectancy between Liverpool, and other areas of the country persist to this day; for example, there is over a 10-year difference in both male and female life expectancy at birth comparing a child born in Kensington in Liverpool (74.8 years for men, 79.2 for women), to a child born in Kensington and Chelsea in London (85.1 years for men, 89.8 for women) (ONS, 2011b). Furthermore, between small areas within Liverpool there are similar differences in life expectancy, of over 10 years, comparing the least to the most deprived areas (DH, 2012a).

We therefore begin with a number of propositions about the inequalities in life chances that are evident within countries wherever they are studied, whereby the most disadvantaged in society have poorer health outcomes. First, these differences are not ‘natural’, or biologically pre-determined, since there is no plausible genetic basis for such large differences, and they largely result from exposure to adverse living conditions. Second these differences are unfair, because they are amenable to public health action, and thus to a large extent preventable (Whitehead, 1990). Third, addressing these differences is a key focus for public health efforts and a matter of social justice, since they constitute a loss of life “on a grand scale” (CSDH, 2008); for example, a recent study of 30 member countries of the Organisation for Economic Co-operation and Development (OECD) estimated that around 10% of the adult mortality in the 15 to 60 age group in these countries, accounting for 1.5 million deaths annually, could be avoided by reducing income-related inequalities to the median level (Kondo et al., 2009).

Having set the scene, this literature review first describes some general theories pertaining to health inequalities, before building the case for studying CF in order to better understand pathways to health inequalities. I then move on to the relevant general CF literature, before focussing in on the key studies that describe inequalities in outcomes in CF. I shall begin with a description of the UK policy context with regards to health inequalities, and its focus on the health of children.

## **The policy context to health inequalities**

Reducing inequalities in health is a public health imperative at local, national and international levels, and strategies to reduce health inequalities have been developed in countries around the world (Judge et al., 2005), including most recently a high profile report from the US (Institute of Medicine, 2013), a country that has tended to shy away from explicit recognition of the health inequalities agenda. A major step-forward has been the adoption of a ‘social determinants of health (SDH) approach’ by the global health community, outlined in The World Health Organization (WHO) Commission on Social Determinants of Health in 2008. The report stated that the development of a society, rich or poor, should not only be judged by the overall quality of its population’s health, but also how fairly it is distributed across society (CSDH, 2008).

The UK has traditionally led the way in terms of health inequalities research and policy, a history that can be traced from the Black report (Black et al., 1980), through the Acheson Inquiry (DH, 1998a), to the recent UK Marmot review (Marmot et al., 2010b). Following the recommendations made in the Acheson report, and the election of the Labour Government in 1997 with their democratic mandate to address inequality, the UK became the first European country to systematically implement and evaluate policies aimed specifically at reducing health inequalities (Mackenbach, 2011a). In particular, the government set targets to reduce the gap in life expectancy and Infant Mortality Rate (IMR) by 10% between the fifth of local authorities with the worst health and deprivation indicators (the Spearhead group) and the population as a whole, by 2010 (DH, 1999, DH, 2003).

In order to address these targets a raft of policies was announced, including the introduction of a national minimum wage, higher benefits and pensions, and substantially increased spending on education, housing, urban regeneration, and health care. There was also the introduction of the ‘Health Action Zone’ programme to improve health in deprived areas (DH, 1999). With regard to improving child health, the ‘Every Child Matters’ agenda (Department for Education and Skills, 2004) united a number of policies, including strategies to increase immunisation rates; achieve early antenatal care for disadvantaged groups; reduce obesity (Cross-Government Obesity Unit, 2008); and reduce parental smoking (DH, 1998b). Most

notably, there were also high-profile targets set around poverty and social exclusion, particularly to cut and eventually ‘eradicate’ child poverty, within 10 to 20 years (Blair, 1999). These policies broadly focussed on three main areas (Judge, 2012):

- Changes to the tax and benefits system – notably the tax credit system, which resulted in progressive improvements in the level of in-work and out-of-work incomes for families with children, and which achieved high coverage.
- Promoting parental employment – increasing incomes for those in low-paid work, and making work pay better for parents e.g. policies to increase the provision of universal, high quality pre-school care; to ensure flexible parental leave; and the requirement that local authorities ensure adequate provision of childcare for parents who want to work.
- Policies to improve children’s life chances – these focussed on the Sure Start centres, which aimed to reduce child poverty through the targetted provision of pre-school education as a means of improving early child development, and subsequent school readiness.

Despite these efforts, by 2010, the UK had not met the target for life expectancy. The data for the two targets show substantial improvements in both life expectancy and infant mortality over the period in both the disadvantaged Spearhead areas, and the other local authorities, but the improvement was slightly greater in the more affluent areas for life expectancy (Marmot et al., 2010b), which has led to an increase in the life expectancy gap. With regard to IMR, there was initially a pronounced increase in the gap, which peaked at 19% between 2002 and 2004 but has narrowed subsequently. We await the final data for the infant mortality target (three year rolling average data) later this year, and it is possible this target may have been achieved, if the recent narrowing of the gap has continued.

There has been much soul searching amongst the health inequalities community, in the UK and internationally, in response to the limited success of the concerted UK drive to reduce health inequalities (Bambra et al., 2011, Mackenbach, 2011a). Various explanations have been put forward. The Marmot review suggests that not enough has been done to address the root causes of inequalities in the social

determinants, and that actions have not been sufficiently co-ordinated across sectors (Marmot et al., 2010b). Furthermore, the tendency for ‘lifestyle’ drift has been implicated, whereby policies start off with a broad recognition of the need to take action on the wider social determinants of health but end up focussing on individual lifestyle factors (Whitehead and Popay, 2010). Mackenbach’s diagnosis is that the strategy did not address the most relevant entry-points, did not use effective policies, and was not delivered at a large enough scale for achieving population-wide impacts (Mackenbach, 2011a). Furthermore he suggests that there was not the political will to address income inequalities, which continued to rise over the period. Our recent analysis supports this, and suggests that if inequalities in household income had been tackled, along with further reductions in unemployment in deprived areas, then the life expectancy target could have been met (Barr et al., 2012). Policy makers and researchers will continue to reflect on the UK experience as new policies are developed to reduce inequalities.

A key theme that has emerged across all of these policies, from Black to Marmot, is the importance of investing in child health. The Commission on the Social Determinants of Health’s overarching recommendation, for instance, relates to improving daily living conditions for all and mandates a renewed focus on child health, with major emphasis on early child development (CSDH, 2008). This focus on the early years is echoed in the Marmot review of health inequalities (Marmot et al., 2010b), which remains the blueprint for action on inequalities in child health in the UK, along with the more recent Field (Field, 2010) and Allen reports (Allen, 2011), which focus on the importance of early child development. As Law points out, reducing health inequalities is likely to require sustained action over time, and the effects may not immediately be apparent (Law et al., 2012). Indeed, the reduction in child poverty over the last decade as a result of the UK strategy, one of the few successes (Judge, 2012), has not yet been evaluated for its impact on health. Any health effects as a result of this intervention may take decades to become fully apparent, in terms of improved adult health.

In addition to the social justice argument for addressing health inequalities, there is also a clear economic case. A recent study estimated that in the European Union 20% of health care and 15% of social security costs could be attributed to health inequalities, leading to a 1.4% reduction in gross domestic product (GDP) through

reduced labour productivity (Mackenbach et al., 2011). Furthermore, addressing disadvantage in the early years has been identified as being critical to tackling some of the most costly health and social problems, such as unemployment, adult chronic diseases, poor mental health, and drug addiction, all of which are related to early disadvantage (Field, 2010, Marmot et al., 2010b). For instance, a Joseph Rowntree Foundation report estimated that the additional cost to public services as a result of the consequences of childhood disadvantage in the UK was between £11.6 and £20.7 billion in 2006/07 (Hirsch, 2008). Furthermore, the New Economics Foundation has estimated that the cost to the UK economy of continuing to address current levels of social problems related to childhood disadvantage will amount to almost £4 trillion over a 20 year period. This is an incredible sum; roughly five times the current annual budget of the whole of the NHS. The report suggests that increased investment in a combination of targeted early years interventions and universal childcare and paid parental leave could help address as much as £1.5 trillion worth of this sum, putting the UK on a par with the Scandinavian countries in terms of child wellbeing (Aked et al., 2009).

Maintaining a focus on reducing inequalities in child health is critical if we are to mitigate the long-term effects of the current economic recession on health (Sell et al., 2010). Child poverty levels are set to rise, and recent analyses of current austerity policies in the UK suggest that children are amongst the groups being hit hardest by these changes (Browne, 2012). Furthermore budget cuts to local authorities are heaviest in the most deprived areas, which is likely to further exacerbate health inequalities (Taylor-Robinson and Gosling, 2011). Before discussing the mechanisms that generate health inequalities, the next section reviews some of the evidence on the corrosive effects of early disadvantage, particularly child poverty, on health outcomes.

### **The effects of childhood disadvantage**

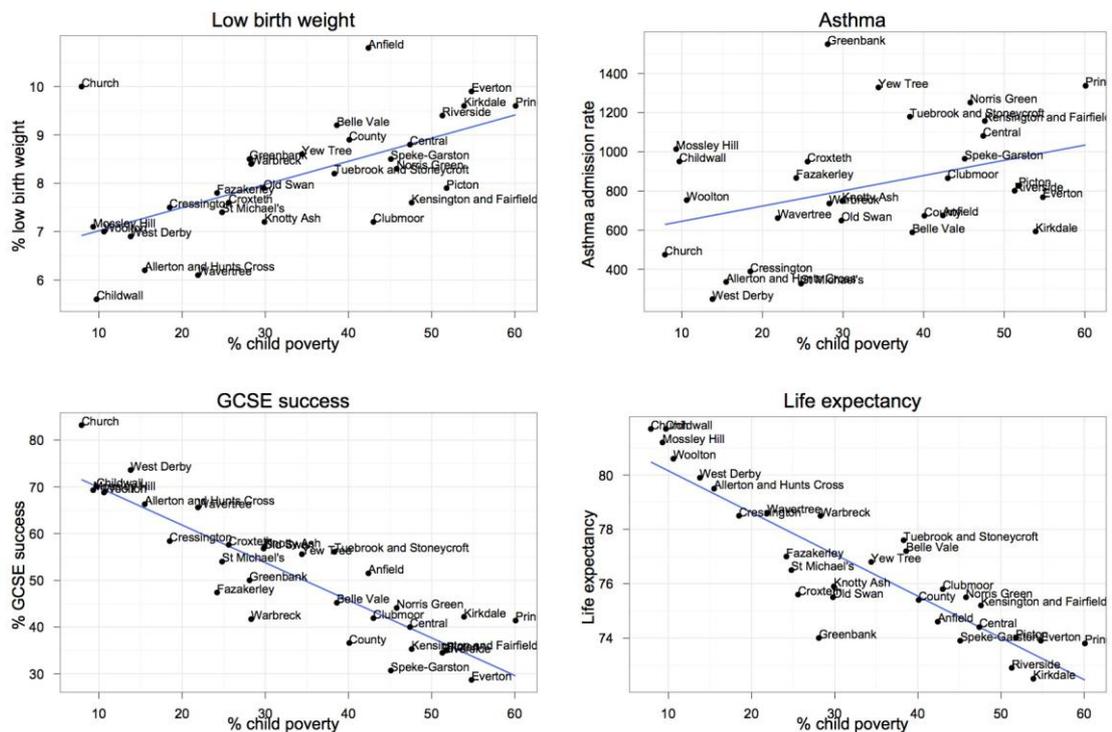
Child health in the UK and other OECD countries continues to improve, with ongoing average declines in IMR, childhood deaths, and improvements in life expectancy. Despite this, unacceptable inequalities persist between rich and poor within the UK. For example, children living in poverty (<60% median income) in the UK are more likely to: die in the first year of life; be born small; be bottle fed; be

exposed to second-hand smoke; become overweight; perform poorly at school; die in an accident; become a young parent; and to die earlier (Roberts, 2012).

In order to provide an illustrative example of the associations between child poverty and health and social outcomes, we turn to contemporary data from Liverpool, one of the most deprived local authorities in the UK. Figure 2 shows the association between relative poverty (<60% median income), and the prevalence of low birth weight; asthma admissions; GCSE success; and life expectancy at ward level in Liverpool.

**Figure 2: Low birth weight; asthma admission rate; GCSE success rate; and life expectancy versus child poverty for wards in Liverpool 2011**

*Asthma admission rate is per 10,000 per year. GCSE success defined as 5 or more C grades or above. Data from Liverpool PCT for 2011, supplemented with child poverty data at ward level from <http://www.endchildpoverty.org.uk/why-end-child-poverty/poverty-in-your-area> (accessed 14th May 2013). Plots generated by author with ordinary least squares (OLS) regression lines in blue*



Children from poorer areas are born small (Figure 2 top left panel), and in longitudinal studies this has been associated with poor growth, cognitive development, and chronic diseases later in life, such as diabetes and hypertension

(Kuh et al., 2004). Maternal smoking in pregnancy is one of the important avoidable causes of low birth weight. In a recent study in Liverpool maternal smoking in early pregnancy was about six times more common in women from the most deprived quintile (based on Index of Multiple Deprivation [IMD]) compared to the least deprived, and was associated with increased rates of preterm birth and low birth weight (Taylor-Robinson et al., 2011). In the first few years of life, children living in deprived areas are also more likely to be exposed to passive smoking (Field, 2010), and are less likely to be breastfed (Kelly and Watt, 2005), illustrating how health-damaging factors begin to cluster and interact at an early stage.

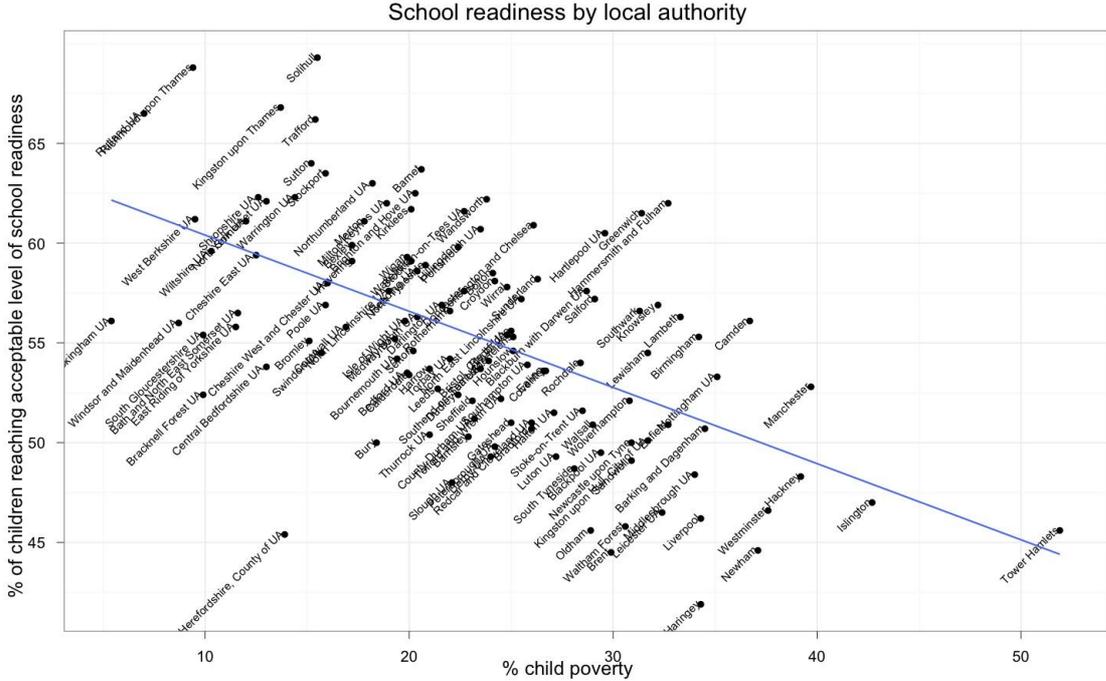
Low birth weight, being bottle-fed, and exposure to second-hand smoke are all associated with a higher risk of developing asthma, which is the most common chronic illness in childhood (Panico et al., 2007). Asthma admissions in Liverpool are accordingly closely associated with levels of child poverty (Figure 2, top right panel). The association between poverty and poor health in the early years extends to many other physical conditions and threats to health in childhood, including chronic ear infection, pneumonia, tooth decay, obesity, and particularly accidents, both fatal and non-fatal (Hirsch D, 2008, Marmot et al., 2010b). Children growing up in poverty are also more likely to suffer a wide range of behavioural and emotional problems: child poverty is related to higher rates of attention deficit hyperactivity disorder (ADHD), bedwetting, self-harm and suicide (Bradshaw, 2011).

The first years of life are crucial for brain development and provide the foundations for children's capacities to learn (Allen, 2011). Furthermore there is good evidence to show that if children fall behind in early cognitive development, they are more likely to fall further behind at subsequent educational stages (Ferguson et al., 2007). Educational trajectories thus track over the course of children's lives, and within wards in Liverpool, the strength of the relationship between child poverty and GCSE success (five or more A to C grades) is striking (Figure 2, bottom left panel). Figure 3 below shows a measure of 'school readiness' for local authorities in England. This is a routinely collected assessment of children's development at the age of five, based on their behaviour and understanding. In order to be school ready, children should be able to share, self-motivate, co-operate and concentrate by the time they start school (Marmot et al., 2010b). The plot shows the stark relationship between levels of child

poverty and school readiness in the UK, and demonstrates that many children in the UK are not achieving their early educational potential.

**Figure 3: Children achieving a good level of development at age 5 versus child poverty**

*Data for upper-tier local authorities in England, 2010, data source London Public Health Observatory, plot generated by author with OLS regression line in blue*



There is great concern amongst policy makers that disadvantage in early life means that children who grow up in poverty are less likely to be in work, to live in a decent home, to earn a decent wage, and to report good health and wellbeing as adults (Field, 2010, Marmot et al., 2010b, Allen, 2011). Thus children who have grown up in poverty are more likely to face recurring poverty and other disadvantages harmful to their health as adults. It is this accumulation of health damaging exposures over the course of people’s lives that is likely to explain the close relationship between levels of child poverty, and life expectancy at birth in Liverpool (Figure 2, bottom right panel).

## **Understanding the causes of health inequalities**

This section focusses on some of the key theories and concepts that have been developed to explain how health inequalities are generated: the social determinants of health (SDH) approach; socio-economic position; material, behavioural, and psychosocial explanations of social causation; health selection; and the life-course perspective. After discussing these issues, we turn to the Diderichsen model for understanding pathways to inequalities in health, which will inform the analyses in this thesis.

### **Social determinants of health**

The SDH approach, adopted by the WHO for the global Commission on Social Determinants of Health, and subsequently the UK Marmot review of health inequalities in the UK, recognises that health inequalities are caused by social inequalities, and arise as a result of differences in living conditions, and the exposures and consequences that follow. The social determinants are the “conditions in which we are born, grow up, work and live” (Marmot et al., 2010b), and include things such as a decent education, adequate housing quality, being able to access a nutritious diet, and having the financial resources to engage fully in society (Taylor-Robinson and Schechter, 2011).

The SDH approach is the dominant paradigm for understanding health inequalities. Its history can be traced back to the likes of Dr Duncan, introduced earlier, and the great social reformers. Engels, for instance, in his studies of the living conditions of the English working class in Manchester vividly described how inequalities in living and working conditions – social determinants – led to ill health in the poor:

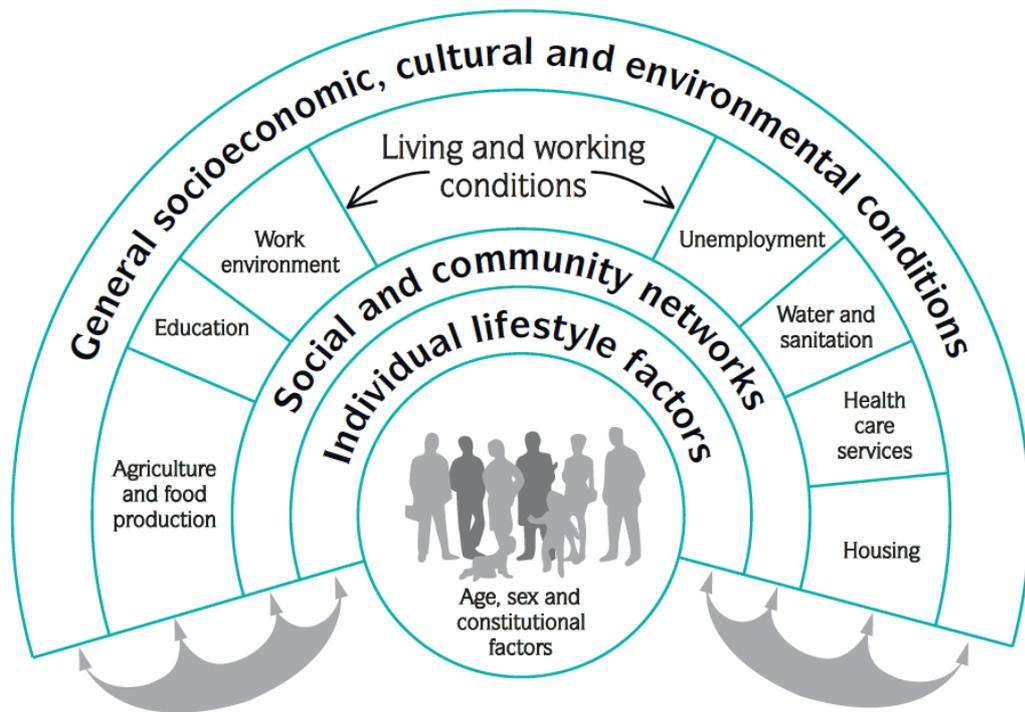
*“All conceivable evils are heaped upon the poor...they are given damp dwellings, cellar dens that are not waterproof from below or garrets that leak from above... They are supplied bad, tattered, or rotten clothing, adulterated and indigestible food. They are exposed to the most exciting changes of mental condition, the most violent vibrations between hope and fear... They are deprived of all enjoyments except sexual indulgence and drunkenness and are worked every day to the point of complete exhaustion of their mental and physical energies.” (Engels, 1845)*

Likewise Thomas McKeown turned to the social determinants in his explanation for the dramatic fall in mortality due to infectious diseases from 1850 onwards in England and Wales. These dramatic changes preceded the introduction of specific clinical interventions such as vaccination, and he hypothesised that the changes were due mainly to improvements in living conditions and nutrition (McKeown, 1979). The focus on structural and material determinants of health can be further traced through the Lalonde report in Canada (Lalonde, 1974), and the WHO Health for All agenda in the late 1970s, which aimed to achieve a decent level of health for all by 2000 (WHO, 1981).

The Dahlgren and Whitehead rainbow is one of the clearest contemporary expressions of the SDH approach (Dahlgren and Whitehead, 1993) (Figure 4).

**Figure 4: The main influences on health**

*Used with permission source (Dahlgren and Whitehead, 2007)*



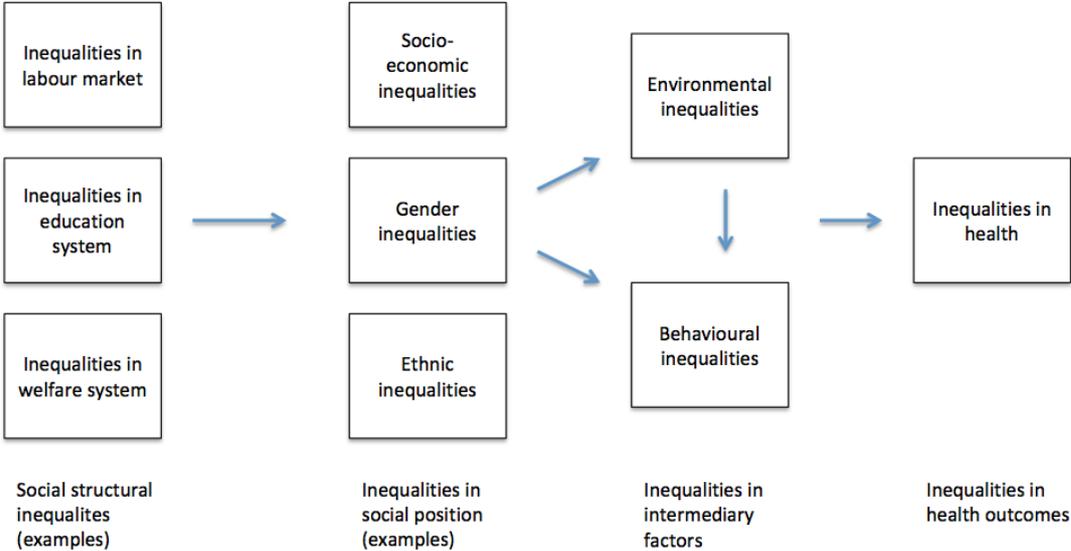
In the rainbow, we see the individual situated at the centre of concentric layers of influence. Thus individual characteristics such as age and one's genetic make-up have an important influence on health status, but we cannot change these. Then there are the behaviours one adopts – what we choose to eat and drink, whether or not we smoke, and who we have sex with for example. While we can change these the model recognises that these factors are influenced in turn by our social and community networks – our immediate surroundings, the neighbourhoods we live in, and the families and friends that exist around us. Then come the leviathans of the SDH approach – the building blocks of health – including the education system, the labour market, systems of food production, and health services. Finally, these factors are in turn influenced by the broader socio-political and physical environment, which can have profound effects on downstream factors in the segments of the rainbow below.

The issue remains, however, as to the underlying mechanisms by which the social determinants exert an influence on health, and a further step required to explain how differential effects arise (Graham, 2004). There has been much conceptual

development in this area, and there are now a number of frameworks that attempt to capture the processes by which the social determinants of health can be linked to inequalities in health outcomes (Mackenbach et al., 1994, Diderichsen, 1998, Whitehead et al., 2000, Brunner and Marmot, 2006). Hilary Graham distils some of the key features of these models in her generic model (Graham, 2007) (Figure 5).

**Figure 5: The social determinants of health inequalities**

*Adapted from Graham, 2007*



In Graham’s model it is the unequal distribution of the ‘upstream’ social determinants of health, which make up the social structure or context, which first lead to inequalities in ‘social position’. Differences in social position between groups and individuals then result in inequalities in exposure to intermediate factors that influence health, and consequently to health inequalities between individuals or groups on the basis of their social position.

**Social position**

There are many dimensions of social position, for instance all of us occupy differential positions in society with regards to gender, culture, ethnicity, sexuality and levels of physical impairment. In this thesis, we are particularly interested in ‘socio-economic position’ (SEP): the position individuals or groups hold within the structure of a society derived on the basis of social and economic factors (Galobardes

et al., 2007). SEP is a complex concept that cannot be considered in any depth here, but many of the theories around SEP can be traced back to those of Marx and Weber (Lynch and Kaplan, 2000). Marx's focus was on social class, as defined by degree of control over the means of production (e.g. land or factories), whereas the Weberian approach suggests that societies naturally become categorised into hierarchical structures, along many dimensions, leading to shared experiences and life chances in particular groups (Galobardes et al., 2006a). Several measures of SEP are used within social epidemiology, and these can be applied at the individual or aggregate level (Galobardes et al., 2006b), but Kunst and Mackenbach have argued that the most important of these are occupational status, level of education and income level (Mackenbach et al., 1997, WHO, 2007).

### **Social causation**

The mechanisms by which unequal social position leads to inequalities in health outcome – the process of ‘social causation’ – need unpacking further. Hilary Graham's model suggests that the effect on health inequalities is indirectly mediated, through a range of more specific intermediate influences that promote or harm health, which are differentially distributed across socio-economic groups. The unequal distribution of these factors subsequently induces higher or lower prevalence of health problems. In terms of understanding which intermediate factors are important in this process, there are more theories to contend with, hotly debated in social epidemiology in terms of their relative importance (Judge, 1995, Lynch et al., 2000, Marmot and Wilkinson, 2001). These relate to material, behavioural or cultural, and psychosocial explanations.

*Material explanations* focus on the relationship between SEP and unequal access to tangible resources at various levels, and was the explanation favoured by the authors of the Black report (Black et al., 1980, Blane et al., 1997). These include, for example, at the individual level, inequalities in access to resources for health, such as food, clothes, and decent housing; and at the aggregate level, sanitary environmental conditions, and safe neighbourhoods and working environments. Beyond the bare necessities for survival it is also necessary to consider the resources required to live a full life in contemporary society (Morris et al., 2000) – a mobile phone, home computer, and a holiday from time to time perhaps – these may also be important in

determining inequalities in health outcomes. Neo-materialist theories are related, but focus on the relationship between resources spent at an aggregate level on factors such as infrastructure, welfare, and the education system, and the magnitude of health inequality (Mackenbach, 2012).

*Behavioural explanations* focus on the distribution of health behaviours such as smoking and alcohol consumption, by social position (Blaxter, 1990). We can take this in a number of directions. Bartley suggests that implicit in the direct behaviour model is the suggestion that people in less advantaged positions are more likely to adopt unhealthy behaviours because they are less endowed with some personal characteristics, read as intelligence, education, self-control, or resilience (Bartley, 2004). The cultural explanation, on the other hand, sees particular behavioural patterns as being “embedded within the social structure” (Black et al., 1980).

The Whitehall studies made use of the finely graded social hierarchy in the British civil service to explore inequalities in health outcomes. Marmot et al. showed that the social gradient in cardiovascular disease outcomes amongst civil servants was only partly explained by behavioural risk factors, suggesting that there were other mechanisms at play (Marmot et al., 1984, Marmot et al., 1991, Marmot et al., 1997). This has led on to *psychosocial explanations* for health inequalities that focus on the role of stress, mediated by neuro-endocrine pathways, caused by living under conditions of relative social disadvantage. For example, increased risk of exposure to negative life events, job strain, stressful living circumstances, high debt, and lack of social support associated with low SES may have direct effects on physiological systems. Stress may also affect health indirectly by leading to harmful health behaviours such as smoking and excess drinking (Kawachi et al., 2002). Furthermore, the Wilkinson hypothesis suggests that the stress caused by constantly comparing oneself with other people in unequal societies is an important driver of health inequalities (Wilkinson, 2005, Wilkinson and Pickett, 2010). Based upon ecological country level analyses in OECD countries and individual states in the US, Wilkinson suggests this can influence overall population health, making more unequal societies unhealthier overall. This approach has been used to relate income inequality to poor health (Wilkinson, 1992), including measures of child wellbeing in the UK and the US (Pickett and Wilkinson, 2007).

## **Health selection**

In contrast to the theories of social causation described above, the selection perspective assumes that health determines SEP, instead of SEP being the starting point, leading on to differences in health. The basis of the selection hypothesis is that health influences social mobility, both between generations (inter-generational), and over the course of people's lives (intra-generational). The health selection perspective suggests that people who are in poor health drift down the social scale, whilst the healthy move up. 'Direct selection' suggests that health directly influences social trajectories, such that sickly children are more likely to find themselves in a lower social position as adults. Indirect selection is more nuanced, and suggests that health status is correlated with other behaviours and attributes (e.g. intelligence, personality, or coping style) that are then associated with social mobility. This thus implies that social mobility is associated with determinants of health, not health itself (West, 1991, Goldman, 2001).

Two contrasting questions arise with regard to health selection. The first consideration is whether health status influences subsequent social mobility and social outcomes, and the second consideration is to determine any effects of selection on inequalities in health outcomes. There is empirical evidence from the UK (Power et al., 1996, Manor et al., 2003), Sweden (Lundberg, 1991) and the Netherlands (van de Mheen et al., 1998) suggesting that illness in childhood does reduce intra-generational social mobility but these effects were not evident in adulthood – for instance, there was little evidence that poor health influences social mobility in the cohort of civil servants in the Whitehall cohort (Chandola et al., 2003).

In terms of the effect of health selection on health inequalities, the results have been mixed, with the majority of studies suggesting that that health selection may actually decrease health differentials (Power et al., 1996, Bartley and Plewis, 1997, Blane et al., 1999, Claussen et al., 2005), although some studies have suggested the opposite (Dahl, 1993, Manor et al., 2003). The mechanism for the reduction in health inequalities as a result of selection is that people who are downwardly mobile because of their health still have relatively better health than the people in the class of destination (Bartley 1997).

Marmot et al offer a robust rebuttal of the health selection approach (Marmot et al.,

2010a), in response to a critique of the Marmot review by the economists Chandra and Vogl (Chandra and Vogl, 2010). Chandra et al suggest that many of the associations between measures of income and health cited in the review, can essentially be explained by health selection (Chandra and Vogl, 2010). In response, after appealing to the writing of Dickens, Marmot et al ask (Marmot et al., 2010a):

*“Should we really assume, that these dark satanic mills and airless places, rather than causing terrible illness and shortened lives, selectively employed sick people and those whose backgrounds accounted for all their subsequent illness? That subsequent improvement in living and working conditions, thus abating Victorian squalor, and associated improvements in health were correlation not causation? That while medical care improved health, housing also got better, and an intellectually slack public health profession mistook the improvement in housing and working conditions for causes of improved health?”*

### **Genetic explanations**

From time to time a genetic argument for health inequalities has been put forward (Davenport, 1916, Herrnstein, 1971 , Himsforth, 1984). Here genetic factors responsible for damaging health are hypothesized to be more common in people with lower SEP. The argument suggests that if a particular gene causes ill health, and people who are ill move down the social strata as a result of the selection process, then this could plausibly lead to the social distribution of genetic material responsible for causing illness, to the disadvantage of those further down the social strata. Furthermore, genes responsible for health promoting characteristics (e.g. cognitive ability, coping styles, control beliefs, personality, bodily and mental fitness) may move up the social scale, and influence inter-generational social mobility.

Such arguments are difficult to sustain. Holtzman highlights key problems with genetic explanations, pointing out that single genes cannot explain complex social traits, and that the timescales required for selection effects cannot explain the more rapid socio-economic changes in health that have been observed in many countries. He concludes, as did the Black report, that the “roots of social class differences do

not lie in our genes” (Holtzman, 2002). Mackenbach further highlights the lack of empirical data to support the genetic argument (Mackenbach, 2005). One of the clear problems with a focus on genetic or biomedical focus on health inequalities is that these explanations can be used as a smoke-screen for maintaining a particular ruling class and reducing social support for the poor, and can distract attention from determinants of health that are policy sensitive.

### **Life-course approaches**

The life course perspective provides a framework for combining all of the aforementioned pathways, by considering the timing and ordering of exposures and outcomes, within an individual’s life course, and across generations. The approach hypothesises that health-protective and health-damaging influences may be more or less important at particular times in life, and that these influences may have effects that accumulate, and interact over time. For example, children from more disadvantaged backgrounds are more likely to be born small, less likely to be breast fed and to have poorer diets subsequently, to be more exposed to passive smoking and some infectious agents, and to have fewer educational opportunities (Ben-Shlomo and Kuh, 2002). Each of these exposures influences and interacts with the risk of subsequent exposures, to influence both health and social outcomes in later life (Galobardes et al., 2004, Kuh et al., 2004, Galobardes et al., 2008).

This approach has played a key role in shaping our understanding of how factors in early life can influence subsequent health and social trajectories. A person’s SEP is not fixed, and can be considered as a trajectory that moves from parental SEP – the class into which one is born – to one’s final adult SEP, measured on the basis of education, income, or employment status for example. In parallel to this trajectory, there is the process of health and capacity development, and eventual decline over time. The life course perspective considers the complex interaction between these two processes, allowing for both social causation and selection over time, and longitudinal studies using individual level data have been the key to developing the evidence base.

A recent systematic review of life course studies reviewed the effects of SEP in childhood on mortality in adulthood. Across 29 studies, the review found that children from less advantaged backgrounds had a higher risk of death in adulthood

across almost all conditions studied, including mortality from stomach cancer, lung cancer, haemorrhagic stroke, coronary heart disease, and respiratory-related deaths, accidents, and alcohol-related causes of death (Galobardes et al., 2004, Galobardes et al., 2008). These studies demonstrate that the conditions in which children grow up are not just critical for child health, they are also critical for adult health, and this has profound public policy implications.

David Barker hypothesized that adult chronic disease, particularly cardiovascular disease, had its origins in the perinatal period, and focussed on the pathway from maternal under-nutrition, to low birth weight in children, and the subsequent association with higher rates of hypertension, type 2 diabetes and risk of cardiovascular disease (CVD) (Barker, 1997). Subsequent longitudinal studies have corroborated some of these associations, and there is now reasonable evidence to suggest that CVD risk can be traced back to the early years of life (Lawlor et al., 2004) (Eriksson et al., 2001). Early childhood is a period of dramatic change and development, in terms of physical growth, cognitive development, and emotional and behavioural learning. Environmental exposures, and the effects of parenting and nurture (Waylen et al., 2008) are likely to be particularly important in this period, and the process of biological ‘embodiment’ of health risks and ‘plasticity’ have been described, whereby early experiences mould the growing child, actually influencing how genes are expressed (Halfon and Hochstein, 2002).

Two approaches have been suggested in terms of the timing of exposures. The critical period model suggests that exposures acting at a specific time point can cause structural changes that have long lasting effect on health. Examples include the effects of thalidomide exposure on limb development, lead exposure on neurodevelopment, or to hepatitis B, where risk of liver cancer later in life is linked to early infection (Galobardes et al., 2004). However, the recognition of early-life influences on chronic diseases does not imply deterministic processes that negate the utility of later-life intervention, since the cumulative model suggests that effects later on in life may modify earlier exposures (WHO, 2007). This has been most extensively studied with respect to CVD risk. For example being overweight as an adult increases CVD risk more in people who were disadvantaged and underweight at birth (Rich-Edwards et al., 2005).

We can also observe the interaction of risks over the life course with regard to health behaviours. For example, childhood social circumstances influence risk of smoking in adulthood (Jefferis et al., 2004). Furthermore, these early risks are compounded by subsequent exposures. Hilary Graham's elegant study using the Southampton Women's survey studied the effect of longer term "biographies of disadvantage" on adult smoking status. Eighteen percent of women with no exposure to disadvantage smoked, compared to 36% of those who were disadvantaged as children. However, 63% of women smoked if in addition they left school early, became young mothers, and had other indicators of adult disadvantage.

The pathway from disadvantage, through low birth weight and early adverse influences on nutritional status, to impaired cognitive development is considered particularly important in understanding these long-range effects. For instance there is a graded relationship between birth weight and parental occupational class, measured by national statistics socio-economic classification (NS-SEC) in the contemporary UK population in the Millennium Cohort Study (MCS) (Dex and Joshi, 2004). Furthermore increased birth weight has been associated with improved cognitive ability at subsequent ages in a range of studies (Machin and Vignoles, 2004, Waldfogel and Washbrook, 2010). A particular concern in the UK at the moment relates to the association between childhood disadvantage and 'school-readiness' (as in Figure 3). For example in the MCS, children born to parents from the most disadvantaged income quintile are almost a year behind those in the middle income quintile in terms of cognitive development, even before they enter the education system at age five (Waldfogel and Washbrook, 2010). The life course perspective suggests that this disadvantage is likely to track forward, and to influence adult health in later life. It also emphasises that exposures early in life are involved in initiating disease processes prior to clinical manifestations, and that differences in childhood exposure profiles can influence the timing of the onset of the symptoms of adult chronic disease (Halfon and Hochstein, 2002).

## **The Diderichsen model**

The Diderichsen model (Diderichsen et al., 2001) captures key concepts in our contemporary understanding of how health inequalities are generated, and was used as the conceptual basis for the work of the WHO Commission for Social Determinants of Health (CSDH, 2008). The framework conceptualises the generation of health inequalities occurring through four main pathways (Figure 6): through social stratification itself; because social stratification leads to differential exposure to risk factors; differential vulnerability at the same level of exposure; and differential consequences of ill health. A key feature of the Diderichsen model is that it incorporates both social causation and social selection mechanisms within a common framework, across the life-course. Importantly, the model also explicitly links the broader social environment to the causal pathway at the individual level and makes it clear that the social determinants of health are policy sensitive.

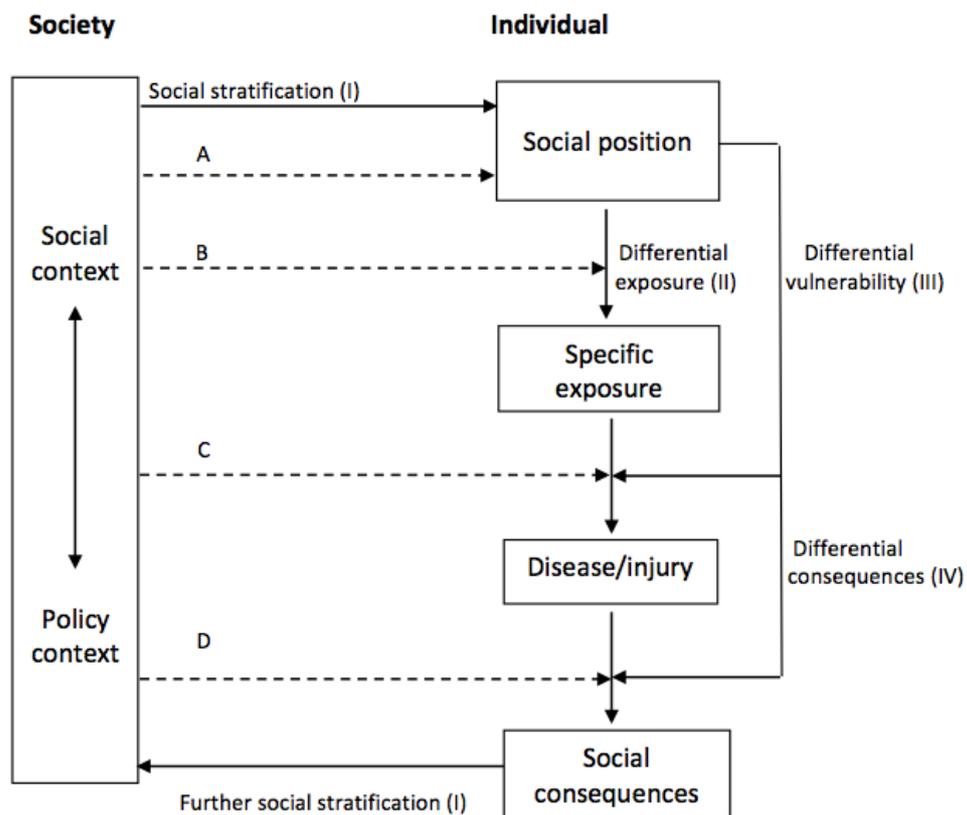
*Mechanism I: social stratification.* The prime mover in the model is the process by which unequal structures in society lead to unequal social positions. Thus the model captures the ‘social causation’ perspective, whereby health is an outcome of processes that begin with social structure in which social position is embedded. The societal structures in question include "those central engines of society that generate and distribute power, wealth and risks" (Diderichsen et al., 2001). So, for example, social inequalities in the education system, the labour market, and the distribution of wealth and resources, first lead to social stratification, as measured by differences in years of education, occupational social class, or income.

*Mechanism II: Differential exposure.* Social stratification influences the degree to which individuals experience health damaging, and conversely health promoting exposures, since health risks vary between social groups in type, amount, and duration. Social position thus mediates exposure to intermediate material and behavioural risks. These include those distributed in the social and material environment of the home, workplace, or neighbourhood. In the absence of any other mechanism, differential exposure may go some way to explaining the differential health outcomes experienced by these groups. For example, children living in poverty are more likely to be exposed to second-hand smoke in the home, and this

may influence their risk of developing respiratory illnesses such as asthma (Panico et al., 2007) and bronchiolitis (Semple et al., 2011). Furthermore, health risks cluster by SES (Halonen et al., 2012), so children living in deprived circumstances are more likely to be exposed to multiple risk factors for respiratory illness, such as smoke exposure, not having been breast fed, and damp housing conditions (Aligne et al., 2000, Almqvist et al., 2005). The life-course perspective recognizes that these exposures are also temporally structured, and occur at every level of development, from the intra-uterine period, through early childhood, childhood, adolescence, and into adulthood. These exposures can thus influence immediate health, and provide the basis for health or illness later in life (Ben-Shlomo and Kuh, 2002).

**Figure 6: A conceptual model for studying the effect of social position and social context on health**

*Adapted from Diderichsen et al., 2001*



**Policy entry points:**

A=modifying effect of social context and policy on social stratification;

B=policies affecting differential exposure;

C=policies affecting differential vulnerability;

D=policies affecting differential social consequences of disease.

*Mechanism III: Differential vulnerability.* This is perhaps the subtlest of the mechanisms, which hypothesises that even if a risk factor is evenly distributed by SEP, it can result in differential effects on health and can be understood in two different ways. Firstly, we can consider a biological mechanism, whereby a person in a lower social position is more vulnerable to a particular exposure. For example, children living in poverty are more likely to become ill or die following exposure to respiratory infections (Bor and Epstein, 1991), and other infections such as malaria or measles, because of the long-term effects of poverty on their host defences, mediated by factors such as malnutrition (Kaler, 2008).

Secondly, we can understand this mechanism as the process of effect modification of exposures, whereby the effect of an exposure is amplified in a particular group, due to the differential level of another risk factor in that group. This is exemplified in risk of cardiovascular disease in adults. The early Framingham studies demonstrated that CVD risk factors modify one another, such that the risk of developing CVD is multiplicative in adults who smoke, have high cholesterol and high blood pressure (Theodorson, 1995). Thus high blood pressure is more harmful to health in individuals who smoke, and it follows that even if high blood pressure was evenly socially distributed, it could lead to differential effects by SEP because of the higher prevalence of smoking in less advantaged groups (Capewell and Graham, 2010).

*Mechanism IV: Differential consequences.* Ill health, as a result of the mechanisms above, can lead to adverse social consequences, such as job loss, or dropout from the education system. For instance, children with type 1 diabetes have been shown to experience some disadvantage in later adult employment chances (Milton et al., 2006). Furthermore, these consequences may be socially patterned, such that more disadvantaged groups are more likely to experience adverse consequences. For example, a study of people with epilepsy in the UK suggested more employment disadvantage in people from working class backgrounds, compared to their more advantaged counterparts (Scambler and Hopkins, 1980). Subsequent longitudinal studies in Sweden have demonstrated that people from less advantaged groups who experience cardiovascular events are more likely to become unemployed than individuals in more advantaged positions (Holland et al., 2009). This final mechanism completes the loop back to the social context, and thus incorporates the

theory of social selection, whereby health status influences SEP, resulting in a pattern of social mobility through which unhealthy individuals may drift down the social scale and the healthy move up.

A great strength of the Diderichsen model is the identification of policy entry points. Thus the model includes a 'social and policy context' box that covers the full length of the individual causal pathway, demonstrating where policies might exert their influence. Policies to influence the process of social stratification (entry point A) would include labour market and welfare policies designed to reduce the gulf between people in different social positions. Policies to reduce exposures (entry point B) include the classic public health interventions such as improvements to housing, safer working conditions, sanitation, clean water, and universal health services. These may be applied universally across a population, but they have the greatest effect on reducing exposure in the worst off, because these groups are the ones suffering the worst conditions (e.g. improvements to working conditions to make them safer will benefit the workers in the unsafe working conditions the most, and the introduction of the NHS benefitted those who had poor access to health services previously). Policies to address differential vulnerability (entry point C) would include strategies that aim to address the multiplicative effect of multiple clustered risk factors, such as co-ordinated population level action on cardiovascular risk factors (Capewell and Graham, 2010). Finally strategies to reduce differential social consequences (entry point D), might focus on provision of health care and social safety nets in proportion to need, in an attempt to reduce any further social differentiation (Diderichsen et al., 2001).

### **Studying tracer groups to better understand pathways to health inequalities**

Studying unique cases, or populations with specific characteristics that allow a new perspective on an issue, has led to important insights in epidemiology and public health. A classic example is the work undertaken by Doll and Peto to study the effects of cigarette smoking in doctors. The fact that doctors were easy to identify, follow-up, and as a whole reduced their cigarette consumption substantially during the period of observation led to some of the first irrefutable evidence that smoking harmed health (Doll 1976).

With regard to health inequalities, the Whitehall studies made use of the finely graded social hierarchy in the British civil service to explore inequalities in health outcomes, particularly CVD (Marmot et al., 1978, Marmot et al., 1991). Similarly, Whitehead et al have used certain groups in the population as ‘tracers’ - sensitive barometers to the prevailing policy and social conditions – to track the effects of policies on health inequalities. The idea is that some groups in the population provide an ‘early warning’ of the impact of changes in policies or services, if the groups most reliant on particular policies are carefully selected for study. Single mothers, for example, have been considered as a ‘litmus test’ for family policy components of the social welfare system because they are among the first to be affected by any changes to the system (Whitehead, Burström & Diderichsen, 2000; Burström et al., 2010).

In a similar manner, we can use the experience of people with the *same* chronic illness as a ‘tracer’ to shed light on the social causes and effects of illness. For instance this approach has been used to examine a number of conditions across the life course such as low birth weight (Pharoah et al., 2003), asthma (Milton et al., 2004), diabetes (Milton et al., 2006), musculo-skeletal conditions (Holland et al., 2006) and end of life care (Hanratty et al., 2007a). Documenting and then analysing the social and health care influences on health inequalities for these groups allows policy recommendations to be made to mitigate the adverse outcomes usually experienced by the most disadvantaged in society. Accordingly, the studies in this thesis utilize the unique characteristics of CF. These hinge on the recessive genetic inheritance pattern, which produces a ‘level playing field’ at the outset, and the ability to explore variation in a range of outcomes across the life course.

### **CF as a tracer condition**

CF is the commonest serious inherited disease among white populations (Davies et al., 2007). Intensive support from family and health care services is needed from the time of diagnosis onwards, and most patients die prematurely from their disease through respiratory failure. There is, however, a great deal of variation in disease progression and survival in CF, that is not explained by genetic differences (Schechter, 2004). This is against a background of astounding improvements in survival over successive birth cohorts in CF. Survival beyond the first few years of life was rare in the 1940s, but UK children born in the 21<sup>st</sup> century are now estimated

to have a median survival of over 50 years of age (Dodge et al., 2007). Survival in CF has increased under the influence of improved treatment and management, improved nutrition and better living conditions (Schechter, 2004, Davies et al., 2007), but these improvements do not appear to have been shared equally, since studies in both the US and UK have demonstrated differences in survival by SES (Britton, 1989, Schechter et al., 2001, O'Connor et al., 2003). The social patterning of survival in CF suggests that social factors – the ‘social determinants of health’ – are having an important effect on outcomes. Utilising some of the unique characteristics of CF to explore the causes of these differences can offer broader insights.

The rationale for studying CF in a public health PhD could be questioned. It is not a common disease of major public health importance at a population level, and does not appear in the global burden of diseases ranking for mortality or morbidity, for instance (Lozano et al., 2012, Murray et al., 2012). Furthermore CF is not identified as being an illness responsible for any significant proportion of the health inequalities ‘gap’ in mortality the UK, which is dominated by CVD, and other chronic diseases of adulthood (SEPHO, 2012).

There are two main responses to this. Firstly, and most importantly, any social differences in CF outcomes are clearly of importance to people with CF, their families, and the clinicians and health services that support people with CF. On this basis alone, it is an important research area. Secondly, and this is what we will focus on here, CF is an interesting case conceptually for the study of pathways to inequalities in health: As a disease of autosomal recessive inheritance where carriers are unaffected, a socio-economic bias in disease incidence is not expected (Schechter et al., 2001, O'Connor et al., 2003), but there is potential for a social gradient in health care use, disease outcome, and social consequences to develop. CF thus opens up opportunities for inequalities research, because it offers both homogeneity, in terms of the population affected by the disease at the outset, and yet great variation in outcomes that can be explored. These two characteristics make for a promising tracer condition to improve our understanding of pathways to differential outcomes.

The genetic inheritance pattern of CF has a number of implications that we can utilize. Firstly, there is no reason to expect a socio-economic gradient in risk of

having CF in terms of incidence (Schechter et al., 2001, O'Connor et al., 2003). From an etiological perspective, we can subsequently follow up the CF population to explore the prospective effects of SES on outcomes, from an equal base. Secondly, cases of CF occur in children whose parents are unaffected by the disease themselves, because of the asymptomatic carrier state. Although more women with CF are living to reproductive age, and having children, genetic screening of partners of women with CF, and antenatal screening in pregnancy means women with CF giving birth to children with CF is a very rare occurrence. Furthermore, the majority of men with CF are infertile (Rowe et al., 2005). Thus from a health inequalities perspective, CF is a condition where the inter-generational transfer of ill health is not a complication in the analysis. Thirdly, because CF primarily affects white populations, the population is relatively homogeneous with regard to race/ethnicity, which are notoriously difficult to separate from SES when considering the causes of health inequalities, since they are intimately linked (Lorant and Bhopal, 2011). Finally the putative even distribution of CF cases by SES is helpful from a study design perspective, and potentially offers a better opportunity to explore differences in outcomes across the social gradient. This is in contrast to diseases where the incidence is socio-economically graded, where there are diminishing numbers of cases in more affluent social groups.

CF allows variation in a range of clinical and health care outcomes to be explored over time, since for example nutrition is particularly important in the early years, whereas lung function is a key outcome in childhood and young adulthood. Despite its origins, explained by a single gene, there is significant variation in outcomes between individuals over time (Schechter, 2011), within individuals over time (Taylor-Robinson et al., 2012a), between clinical care centres (Johnson et al., 2003), and between national populations (Fogarty et al., 2000). This variation is a prerequisite to epidemiological studies. There have also been dramatic changes in survival over time, leading to significant cohort effects (Dodge et al., 2007), and it is possible that patterns of inequalities in outcomes may therefore change over time. There are parallels here with the broader health inequalities context, where one particular challenge has been to explain the persistence of health inequalities, despite the on-going improvements in average health across all groups (Graham, 2007).

Studying CF opens up opportunities to explore all of the major aforementioned theories regarding the generation of health inequalities. Considering material theories of social causation, CF is a respiratory disease where there are plausible links to the effects of housing and local environmental quality on disease outcomes, which may in turn be socially patterned (Britton, 1989). Behavioural factors, such as smoking, and exposure to environmental tobacco smoke (ETS) have been shown to influence CF outcomes (Smyth et al., 1994). With regard to psychosocial theories, stress pathways may be particularly important in the context of a family dealing with the burdens associated with caring for a child with a complex chronic illness like CF, over long periods of time (Taylor-Robinson and Schechter, 2011).

Inequalities in health services are likely to be particularly important determinants of health inequalities in CF, and a key consideration is the role of health services in mitigating or potentiating differential outcomes. Investigating the existence of the ‘inverse care law’, by which the availability of good medical care tends to vary inversely with the need for it (Hart, 1971), is a research priority in CF care in the UK (Taylor-Robinson and Schechter, 2011). Evidence from other diseases suggests that there is a lag time associated with accessing new treatments, such that more affluent groups benefit first (Lyraatzopoulos et al., 2011). This may be important in CF, where new treatments are rapidly emerging.

Investigating these factors from a life-course perspective is now possible thanks to the CF registries in many countries, allowing access to high quality longitudinal data on children with CF, captured from birth, across multiple outcomes. This allows the opportunity to separate age and cohort effects, and to more robustly demonstrate causal associations; it could also allow the exploration of cumulative effects of exposure across the life course. Since CF is severe, life-limiting illness, there is the potential to observe for differential effects over short time periods. Furthermore, CF is an illness where health selection is also plausible, since illness in childhood can affect learning and employment opportunities leading to subsequent effects on SEP.

Having outlined the case for studying CF as a tracer condition to better understand pathways to health inequalities, the next section reviews some key aspects of CF in order to provide sufficient background and context to understand the analyses in this thesis.

## **Key features of CF**

CF is an inherited, chronic, progressive condition occurring in around 1 in 2500 live births in the UK, and affecting over 9000 individuals, with around 300 new diagnoses annually (Dodge et al., 2007, CF Trust, 2013a). CF is the commonest serious inherited disease among white populations, and is caused by mutations in the CF transmembrane conductance regulator (CFTR) gene, resulting in thick secretions that impair various organs, particular the respiratory and digestive systems. It is an autosomal recessive condition; inheritance requires receiving two copies of the defective CFTR gene, one from each parent (Davies et al., 2007). Most children are diagnosed in the first few months of life, and subsequently require intensive support from family and health care services. Most patients die prematurely from their disease through respiratory failure, and in the 1930s and 40s survival beyond childhood was rare. There have been impressive improvements in survival over subsequent decades, such that the estimated median survival of a child born today with CF is over 50 years in the UK (Dodge et al., 2007).

## **Basic defect**

CF is characterised by a generalized defect of all secretory glands in the body (exocrine function), resulting in the production of unusually thick secretions, which cause blockage and subsequent damage to the organs in question, particularly the lungs, and pancreas. Thus, CF used to be primarily a digestive and lung disease of young children but with increasing survival into adulthood, we can better understand it as a complex, multisystem disease.

The thick secretions are the results of the defective opening of ion channels at the luminal surface of epithelial cell membranes, which leads to increased sodium and water absorption from the lumen. The failure of this liquid volume regulation on epithelial surfaces relates to a basic genetic defect in the gene that codes for the ion channel – the CFTR gene, which is located on the long arm of chromosome 7 (Rowe et al., 2005). Identification of the gene for CF in 1989, was a major breakthrough in understanding the aetiology of CF. The most common mutation is delta F508 – deletion of phenylalanine at position 508 (Kerem et al., 1989, Riordan et al., 1989, Rommens et al., 1989), but over 1600 mutations have been recognized and described subsequently (Wolfenden and Schechter, 2009). We can classify these mutations on

the basis of the mechanism by which they are believed to cause disease (class 1 to 6, with higher number being generally less severe). For instance class 1 defects include the complete absence of synthesis of the CFTR, whereas in a class 6 defects a functional CFTR is present at the cell surface, but there is increased turnover, leading to loss of function. Delta F508 is a class 2 defect characterised by defective protein maturation and premature degradation (Rowe et al., 2005).

There is great variation, however, in phenotype in CF, which is not explained by genotype. This means that progression of lung disease, the key mediator of survival, has proven very difficult to predict in individuals. Thus, while the manifestations of lung disease, as well as mortality, are greatest in patients with the more severe class 1-3 mutations (Wolfenden and Schechter, 2009), there is generally a loose genotype-lung phenotype relationship, as demonstrated by significant variation in CF lung disease, even among individuals with identical mutations (Kerem et al., 1990a, 1993, Koch et al., 2001). This has led to a number of recent studies into modifier genes that may influence the effect of the CFTR mutation (Wolfenden and Schechter, 2009, Cutting, 2010), most notably the TGF $\beta$ 1 gene where there seems to be consistent evidence to support its association with poorer lung function (Drumm et al., 2005). It has also led to the recognition that despite having its origin in a single gene defect, a large amount of the variability in clinical course of CF is not determined by genetic factors (Wolfenden and Schechter, 2009, Schechter, 2011). A recent twin study by Collaco et al, for instance, suggested that approximately half of the variation in lung function is attributable to environmental and/or other stochastic (random) factors (Collaco et al., 2010), with the other half being due to genetic factors, affirming the potential role of genetic modifiers that require further investigation.

### **Clinical presentation**

The clinical presentation of CF varies according to age. Genetic screening and counselling are offered to couples with a family history of CF. Prenatal diagnosis can then be undertaken by amniocentesis or chorionic villous sampling from around 10 weeks gestation (Wald et al., 2003). CF may also be suspected on the basis of detecting echogenic bowel on antenatal ultrasound scanning (Scotet et al., 2002).

In the neonatal period (<28 days), diagnosis is usually as result of screening, or due to meconium ileus. All babies born in the UK are offered screening soon after birth for certain rare medical conditions, including CF, which has been universally available throughout the UK since October 2007 (NHS, 2011). The number of children diagnosed through screening has risen year on year in the UK subsequently, such that in 2010, out of 301 newly diagnosed cases of CF, 63% (189) patients were identified by newborn screening (CF Trust, 2013b). Neonatal screening is based on demonstration of low levels of immunoreactive trypsin in the Guthrie heel prick sample of blood (NHS, 2011).

It is hoped that screening will allow more early intervention, before any lung disease is established, and this will lead to long-term benefits for patients with CF (Greasemann and Ratjen, 2013). The Cochrane review of randomised control trials (RCTs) suggested that severe malnutrition was less common among screened babies (Southern et al., 2009). Further support for newborn screening has come from observational registry based studies, which have tended to support the theory that early diagnosis within the first two months of life, followed by intensive treatment, is likely to improve nutrition, and early lung function, but evidence is lacking for a long term effects on survival (Salvatore et al., 2010). Furthermore, there has been a number of cross sectional studies using the UKCF database that have contributed to this body of evidence, comparing the outcomes of individuals screened on the basis of regional programmes instigated in the UK prior to universal coverage, to those diagnosed symptomatically: The first suggested the achievement of a significantly greater median height and a reduction in morbidity in screened patients as compared to controls matched for age and genotype (Sims et al., 2005b), and two subsequent studies suggested newborn screening led to less acute treatment requirements (Sims et al., 2005a), and reduced health care costs (Sims et al., 2007b). These findings were corroborated in another study by the same group, that demonstrated improved height z scores, but no difference in lung function, comparing screened to clinically diagnosed patients (Sims et al., 2007a).

About 15% of neonates with CF present with meconium ileus at birth, caused by thick mucus blocking the small intestine. This is a life threatening condition if left untreated, usually requiring surgical intervention (van der Doef et al., 2011). Distal intestinal obstruction syndrome (DIOS), sometimes referred to as ‘meconium ileus

equivalent', has a similar pathophysiology to meconium ileus and occurs in older children and adults (Smyth, 2005).

In infants and young children CF may be suspected due to recurrent respiratory symptoms (cough, wheeze, and pneumonia), or failure to thrive. About 90% of people with CF in the UK have pancreatic insufficiency (PI) (authors calculation from registry data), with consequent malabsorption of fat and other nutrients, leading to diarrhoea and subsequent poor weight gain in infancy. In older children and adults, respiratory symptoms, sinus disease or infertility, along with a range of less common presentations, may lead to suspicion of CF and further diagnostic testing (Davies et al., 2007).

When CF is clinically suspected then the diagnosis is usually confirmed on the basis of the sweat test, by demonstrating an elevated level of sodium in the sweat. A raised sodium  $>60$  mmol/L is considered diagnostic, and levels above 40 mmol/L are consistent with CF. Though definitive diagnoses are not possible in all cases, the majority of cases can be diagnosed by a combination of clinical features, sweat testing, or identification of CF specific mutations from DNA blood testing for the more common CFTR mutations (Rosenstein and Cutting, 1998)

### **Lung disease in CF**

Lung disease in CF is particularly important, because the vast majority of deaths in CF are due to pulmonary causes – around 97% in one report (Smyth, 2005). Kerem et al's seminal study in 673 patients from Toronto was the first to quantify the relationship between levels of lung function, as measured by forced expiratory volume in 1 second as a percentage of predicted ( $\%FEV_1$ ) and survival, with a doubling of the risk of death within two years for each 10% decrease in  $\%FEV_1$ . For patients with a  $\%FEV_1$  lower than 30% this corresponded to a 50% risk of death within the next two years. They also demonstrated female sex as an independent risk factor for premature death (Kerem et al., 1992).

In terms of the clinical picture, persistent respiratory symptoms in infants (cough and/or wheeze) are often the first indication of CF, and in children, recurrent chest infections are common presenting features. The natural history of CF lung disease is

then characterized by a chronic progression with intermittent episodes of acute worsening of symptoms termed pulmonary exacerbations. Clinical features of these include increased cough and sputum production, shortness of breath, loss of appetite, loss of weight, and a decline in lung function (Goss and Burns, 2007)

Bacterial lower respiratory tract infections, particularly those due to *Staphylococcus aureus*, *Pseudomonas aeruginosa*, *Haemophilus influenzae* and *Burkholderia cepacia* become established within viscid airway secretions in patients with CF and are not effectively eradicated (Rowe et al., 2005). Infection with *P. aeruginosa* can occur very early on in life (Rosenfeld et al., 2012a), and becomes increasingly common over childhood, such that by the age of 18 around 60% of the UK CF population are infected (Taylor-Robinson et al., 2013a). Infection with *P. aeruginosa* is clearly associated with worse outcomes, including survival, lung function, and nutritional status (Kosorok et al., 2001, Emerson et al., 2002, Konstan et al., 2007a, Taylor-Robinson et al., 2012a, Taylor-Robinson et al., 2013a).

Pulmonary inflammation is another major cause of the decline in respiratory function in patients with CF. Recent evidence suggests that progressive pulmonary inflammation and damage in CF occurs very early in life (Sly and Brennan, 2004, Mott et al., 2012), and perhaps prior to the onset of airway infection (Stick, 2009, Ramsey et al., 2012). For instance, airway damage has been identified in infants with CF by four weeks of age (Mott et al., 2012). One of the key questions in current CF care relates to the effectiveness of early intervention, in pre-symptomatic infants and children, in terms of avoiding early lung function damage. Current thinking suggests that the infant and preschool age period (2 to 5 years) could represent a unique window of opportunity to postpone or even prevent the onset of CF lung disease, with presumed consequent effects on long-term survival (Greasemann and Ratjen, 2013). This is a rapidly growing area, and a pre-requisite for clinical trials is the ongoing development of sensitive and reliable outcome measures in the early years – CT imaging, bronchoalveolar lavage (BAL) for early microbiological profiling, and inert-gas washout tests are current options (Greasemann and Ratjen, 2013).

Regardless of the exact aetiology and timing of early lung disease in CF, by the time children reach school age (5 years of age), significant lung damage (bronchiectasis) is found in most patients despite lung function measures within the normal range (de

Jong et al., 2006). There has been much interest over the past couple of decades in quantifying the subsequent changes in lung function from school-age onwards, using %FEV<sub>1</sub> as the most commonly used outcome. One of the first studies was that of Corey et al, in a single centre in Toronto, which was also the first to use modern longitudinal analysis techniques in CF. They demonstrated that pancreatic sufficiency, male gender, and non homozygote status for the delta F508 mutation were associated with a slower rate of pulmonary function decline (Corey et al., 1997). There have been several important studies of risk factors for lung function decline in CF (Salvatore et al., 2012), using national disease registries, that are considered in further detail in a subsequent section.

### **Management of CF**

The management of CF is multidisciplinary in the UK and includes specialist consultant paediatricians or adult physicians; clinical nurse specialists; physiotherapists; dieticians; clinical psychologists; social workers; pharmacists; secretarial support; and database coordinators (CF Trust, 2011). Following a description of the organisation of care in the UK, this section provides an overview of some of the key therapies in CF that are relevant to the studies in this thesis.

### ***Organisation of CF care***

The CF Trust in the UK, and their counterparts in the US, the CF Foundation have developed consensus guidelines (Flume et al., 2007, Flume et al., 2009a, Flume et al., 2009b, Robinson et al., 2009, Flume et al., 2010) and have promoted a network of CF care centres with expertise in the treatment of CF. The majority of people with CF in the UK attend or receive all or some of their care from one of these specialist CF centres, which are staffed by a multidisciplinary team (MDT) with appropriate expertise and training in the management of CF (CF Trust, 2008).

CF care in the UK, and most other high-income countries, is delivered through these specialist clinics, where a critical mass of patients can be treated, and resources and expertise concentrated. Furthermore, clinicians and services need to have an adequate throughput of patients in order to maintain and improve quality of care. This is an established principle for adult conditions such as stroke (Chan et al., 2013) and cardiac services (Walker et al., 2012), and also for specialised paediatric surgery

(Gibbs and Cunningham, 2002, Stringer, 2008, Welke et al., 2008), and cancer services (Knops et al., 2013), where centre care has improved outcomes.

The first evidence for improved outcomes as a result of centre based care in CF came from Denmark (Nielsen et al., 1988), where patients treated in the Copenhagen CF centre fared better in terms of lung function and prevalence of *P. aeruginosa* infection. Similar findings suggesting better pulmonary function, nutrition and survival in patients receiving centre care compared to those receiving local hospital based care have come from subsequent observational studies in the UK (Walters et al., 1994, Mahadeva et al., 1998), Australia (Phelan and Hey, 1984), Denmark (Merelle et al., 2001) and Germany, and the current consensus is that contemporary care for CF should be delivered by a specialist CF centre MDT (Kerem et al., 2005, CF Trust, 2011). Despite this, a recent systematic review of specialized care concluded that for CF, outcomes were not superior in specialized centres compared with other models of care, commenting on the inconsistency and poor quality of the observational studies in CF (Post et al., 2009).

Doull points out that these early studies, which reflect on care over 20 years ago, may not be generalizable to modern CF care, which has evolved considerably over the last two decades (Doull et al., 2012). More recently, a ‘shared care’ model has been increasingly common (CF Trust, 2008); with similar outcomes demonstrated for children who received specialist care compared to those who received shared care (van Koolwijk et al., 2002, Thomas et al., 2008). Shared care means that the overall patient management is co-ordinated by the MDT based at a specialist centre, but some clinic visits are delivered on an out-reach basis in accredited network clinics. This allows patients to be seen closer to home, in their local paediatric units, usually in conjunction with a local paediatrician with a specialist interest in CF, thereby reducing travel times and disturbance to schooling and family life.

A recent analysis of outcomes in South Wales identified three patterns of care (Doull et al., 2012): full centre care, where the child is seen every 6 to 8 weeks (including annual review) at the specialist CF centre; local clinic-based care (shared care), where the child is seen regularly by the local clinic team and has only their annual review performed locally by the visiting consultant from the specialist CF centre; and hybrid care, where the child is usually reviewed at least three times a year (including

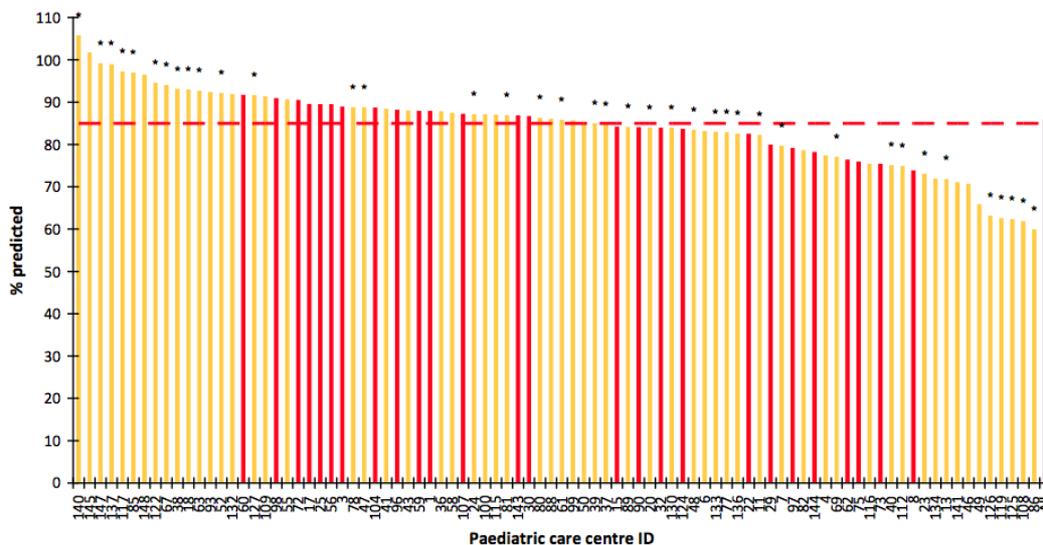
annual review) by the specialist CF centre team (Doull et al., 2012). The study reported worse lung function in the shared care group. The authors suggest that frequency of review by the CF centre MDT is more important than the distance from the CF centre, and hypothesize that the differences observed between models of care may reflect the frequency of specialist CF team review in the first few years of life. Data from the US also suggests that the centres with highest lung function scores for their patients were characterized by more clinic visits, more respiratory tract cultures, and frequent treatment of patients, particularly those considered to have mild lung disease (Johnson et al., 2003, Padman et al., 2007). Notably, the current UK guidelines now recommend that the specialist MDT should see patients under shared care at least twice a year (CF Trust, 2011). It is evident that continuing research is needed to identify the optimum model of care in the UK, and elsewhere, as treatment strategies evolve.

The establishment of clinical networks in the UK and US, taking a standardized multidisciplinary approach to CF care, has contributed to the on-going improvements in survival for people affected by the disease. There remains, however, significant variation in disease outcomes between centres. Data from the US (Schechter, 2012), UK (CF Trust, 2013b) and Germany (Stern et al., 2008) suggest that there is significant variation in outcomes such as lung function and nutritional status comparing centres. Most studies exploring reasons for centre-based variation have been from the US. For instance, registry based analyses using the Epidemiological Study of CF (ESCF), taking into account case-mix, have demonstrated centre based differences in use of therapies and patient monitoring (Konstan et al., 1999b, Konstan et al., 1999a), and have also shown that these variations in care impact directly on outcomes (Johnson et al., 2003). This has led to increased interest in benchmarking and quality improvement initiatives in the US in an attempt to drive up standards (Quinton and O'Connor, 2007, Schechter and Gutierrez, 2010, Quon and Goss, 2011). In a UK analysis, there was a threefold variation between CF clinics in the prevalence of poor nutritional status (<10<sup>th</sup> centile for weight), and Mehta et al suggest that further investigation is required into the factors that might explain such variability, in line with the on-going work in US CF clinics (Mehta et al., 2004). There remains significant variability in all measured unadjusted outcomes between centres in the UK (Figure 7)(CF Trust, 2013b), but no analyses to date have

taken into account case mix, and other factors such as social deprivation, so interpretation is difficult.

**Figure 7: Median FEV<sub>1</sub> % predicted by paediatric centre/clinic all centres and networks.**

*Red: centres. Gold: network clinics. Plum: all. \* Centre/clinic with a dataset submission of less than 20 patients source (Cystic Fibrosis Trust, 2011)*



### ***Respiratory management – antibiotic therapy***

Treatment of respiratory infections with antibiotics is a central pillar of CF therapy including for prophylaxis; eradication of infections; long-term treatment of chronic infection, and treatment of acute exacerbations. The aim is to prevent initial bacterial infection in children, and promptly treat acute infections using antibiotics.

Younger children receive oral antibiotic prophylaxis for *S. aureus* in the UK, but this is not standard everywhere. Current evidence from RCTs suggests that this may prevent infection, but there are concerns relating to increased risk of *P. aeruginosa* acquisition with long term prophylaxis (Smyth and Walters, 2012a).

*P. aeruginosa* is associated with increased morbidity (worse lung function and weight gain) and mortality (Emerson et al., 2002, Taylor-Robinson et al., 2012a), and management aims to prevent early infection. *P. aeruginosa* is widespread in the natural and domestic environment, including plants, soils, and surface water,

especially warm moist environments (Cohen-Cymerknoh et al., 2011). Concerns about cross infection between patients means that complete segregation of CF patients harbouring *P. aeruginosa* and other clinically significant organisms is considered best practice (Jones et al., 2005). There is increasing evidence that early antibiotic therapy initiated early after the initial onset of *P. aeruginosa* infection is effective, and can postpone chronic colonization in over 80% of people, but the long term benefits are unclear (Davies et al., 2007). No consensus currently exists regarding the best antibiotic protocol, and the Cochrane review found that nebulised antibiotics, alone or in combination with oral antibiotics were effective options (Langton Hewer and Smyth, 2009).

Once *P. aeruginosa* infection becomes chronic, long-term treatment with inhaled antibiotics is recommended. The best evidence from RCTs is for inhaled tobramycin, but again the optimum treatment regimen is unclear in terms of dose and frequency of administration (Ryan et al., 2011). Inhaled tobramycin was shown to produce sustained improvements in lung function, improved patient nutritional status, hospitalization time, and the requirement for IV antibiotics (Ryan et al., 2011). Oral macrolide antibiotics have also been shown improve respiratory function, and are used to reduce inflammation in patients with CF colonized with *P. aeruginosa* (Southern et al., 2012). By contrast the Danes have advocated intensive IV antibiotic therapy delivered on a regular basis to improve outcomes in people colonized with *P. aeruginosa*, but this has not been adopted widely elsewhere (Frederiksen et al., 1996).

Acute chest infections (exacerbations) are treated with oral, nebulised, and/or IV antibiotics. It remains unclear whether the optimal treatment should be with oral, inhaled (Ryan et al., 2013), or IV antibiotics (Cohen-Cymerknoh et al., 2011), and whether this treatment should be provided in a hospital setting or at home (Balaguer and Gonzalez de Dios, 2012).

### ***Respiratory – other therapies***

Macrolides, mainly azithromycin, are being used to reduce inflammation in patients with CF colonized with *P. aeruginosa*. High-dose ibuprofen can also slow the progression of lung disease in people with CF, especially in children (Lands and Stanojevic, 2007). Several studies reported beneficial effects of systemic

corticosteroids, but there are risks of significant adverse effects, such as growth retardation, diabetes and cataracts, so these are not recommended (Balfour-Lynn and Welch, 2009).

DNase is an inhaled mucolytic that breaks down the thick sputum in the lungs. This was first commercially available in 1992, and was the first treatment to demonstrate an improvement in lung function in CF (Fuchs et al., 1994). A recent systematic review of subsequent trials demonstrates improvements in lung function from long term DNase therapy (Jones and Wallis, 2010). Inhaled hypertonic saline is also used to improve airway clearance, and has been shown to improve quality of life and reduce pulmonary exacerbations, although it has not been shown to have a substantive effect on lung function (Wark and McDonald, 2009). Inhaled saline has recently been shown to be ineffective in improving outcomes in children under six years of age – it had been hypothesised that this may be an effective early intervention to reduce the rate of pulmonary exacerbations in children (Rosenfeld et al., 2012b).

Daily physiotherapy exercises to improve mucous clearance are recommended for all patients. These include postural drainage, percussion, and vibration techniques, huffing and directed coughing, and can be facilitated with devices such as the flutter or oscillator vests. Although physiotherapy does appear to increase mucous clearance, at present there is no clear evidence from RCTs for long-term effects in respiratory function, quality of life or survival (Van der Schans et al., 2009).

### ***Nutritional support***

The main aims of nutritional support in CF are to achieve optimal nutritional status, and allow normal growth and development throughout childhood. Pancreatic enzyme insufficiency leads to malabsorption of fats, diarrhoea, and failure to thrive, and this is compounded by lung disease and infection, which further increases calorie requirements. Thus recommendations are for early nutritional support with adequate pancreatic replacement management, as this has been shown to improve growth and subsequent lung function (Konstan et al., 2003, Munck et al., 2009). Nutritional status has been independently linked to survival, and the experience of the Toronto CF Clinic suggested that a high calorie diet may improve growth (Levy et al., 1986). However this has not been corroborated in RCTs. The latest Cochrane review

suggests that oral calorie supplements do not confer any additional benefit in the nutritional management of moderately malnourished children with CF over and above the use of dietary advice and monitoring alone (Smyth and Walters, 2012b).

### **Epidemiological studies of CF using national registries**

Registry studies have been critical to inform our understanding of the epidemiology of CF, and I have discussed some of the registry studies relating to newborn screening, and the evidence around centre care above. These observational studies have become increasingly common since the late 60s, due to the availability of registry data in a number of countries. Analysis of registry data has allowed insights into the changing demographics, outcomes, and treatments in CF. A recent series of systematic reviews have collated all of the registry trials in CF, identifying 168 studies overall (Buzzetti et al., 2009, Salvatore et al., 2010, Salvatore et al., 2011, Salvatore et al., 2012). In their analysis of these studies the authors group the studies into the following outcomes: demographics; incidence/prevalence; survival/gender gap; genetics; nutritional status and growth; microbiology; complications; factors influencing diagnosis; factors influencing lung disease; socio-economic status/quality of life; therapeutic strategy evaluation; clinical trial methodology (Salvatore et al., 2012). Eleven CF registries are identified, summarised in Table 1 below.

In this section I provide an overview of registry studies that focus on survival/mortality, nutrition/growth and lung function, mentioning some other non-registry studies where appropriate, since these outcomes are of particular relevance to the studies in this thesis. CF registry studies that focus on the effect of SES are summarised later on. This overview highlights key emergent findings, and evidence from the UK, since this is the main data-source for this thesis, in order to identify key gaps in the literature.

**Table 1: National CF registries**

<b>Registry name</b>	<b>Start date</b>	<b>Patients captured (alive and dead) to end of 2006</b>
USA-CFF Patient registry	1966	40,203
USA-Epidemiologic Study of CF	1994	32,667
Canadian CF Patients Data Registry	1966	5571
Australian CF Data Registry	1988	2312
UK CF Database	1999	7046
France-Observatoire National de la Mucoviscidose	1992	4608
Registro Italiano FC	1988	5064
German CF Quality Assessm. Proj.	1995	7260
Registre Belge de la Mucoviscidose (BMR-RBM)	1998	1140
Cystisk fibrose register Danmark	2001	460
European Registry CF (ERCF)	1995	13,684

***Survival/mortality***

Two reviews identified 19 studies that focussed on changes in survival, and factors influencing survival in CF (Buzzetti et al., 2009, Salvatore et al., 2012), with one study identified from the UK (Dodge et al., 2007). The majority of these were from the US, as a result of the large size and longer follow-up period in the two US registries (Table 1).

The most recent data show a median survival age (indicating the age at which a newborn of a given period has a 50% theoretical chance of surviving) of 36.4 years in France, based on 2003 data (Bellis et al., 2007), 37.4 years in Germany, based on 2005 data (Stern et al., 2008) and 37.4 years in the U.S., based on 2007 data (CFF, 2007). The study from the UK estimates survival at over 50 years for patients born after 2000 (Dodge et al., 2007). All these studies use life-table methods to study survival. The review authors comment that “improved survival, increased number

and increased mean age of CF patients worldwide are evident” (Buzzetti et al., 2009).

Some studies have been able to study factors that influence survival, and notable among these are the large studies from North America that have used Cox regression to estimate the effect of various covariates on survival chances in registry populations (Corey and Farewell, 1996, Rosenfeld et al., 1997, Liou et al., 2001, O'Connor et al., 2002). Overall, the following negative influences on survival have been identified: female sex (the most commonly identified risk factor, see for example (Rosenfeld et al., 1997), one of the largest studies), early symptomatic diagnosis, poor nutritional status, poor respiratory function, diabetes, *P. aeruginosa*, *B. Ceperacia* infection, >4 respiratory exacerbations per year, homozygous delta F508 status or heterozygous non delta F508 status, non-white ethnicity, and low income. Notably, only the effect of sex has been studied in the UK population (Buzzetti et al., 2009).

Turning in more detail to the UK studies, Dodge et al have undertaken a series of epidemiological studies describing survival (Dodge et al., 1993, Dodge et al., 1997, Lewis et al., 1999, Dodge et al., 2007). These have used similar approaches on the same underlying dataset, updated to cover larger follow up periods, and therefore we focus here on the most recent study (Dodge et al., 2007). In order to calculate survival, two main sources of data are required: a complete census of the population and follow-up of that population to capture mortality data (WHO, 1977). Dodge et al used the UK CF Survey (UKCFS), a precursor to the UKCF Registry, which was originally under the auspices of the British Paediatric Association and the British Thoracic Society, and funded by the CF Trust. This enumerated all of the people in the UK population with CF, through direct enquiry with individual clinicians, from 1968 to 1997, with an estimated coverage of over 95% (Dodge et al., 2007). Therefore the dataset captured the prevalent population in 1968, and incident cases born subsequent to this up to 1997. The Office of National Statistics (ONS) collected mortality data by requesting all CF related International classification of disease (ICD) codes, and this was linked back to the survey data. There has been standardised ICD coding for CF since 1968, whereas prior to this ICD did not separate CF from other diseases of the pancreas, so death certification data is unreliable (Dodge et al., 2007). Using this information the authors were able to

calculate cohort survival, and left truncated-survival, and then go on to estimate current survival for the CF population in 2003, on the basis of a number of assumptions.

Cohort survival was calculated from the observed data, based on the survival for incident cases captured subsequent to 1968. This method requires a long follow up period, and estimates of median survival derived from such data are unlikely to be relevant to newly born cases because of improvements in treatment leading to a longer lifespan (Lewis, 1998). Using an actuarial life table method (Armitage et al., 2008), they generate survival curves and age/sex specific mortality rates for birth cohorts in three-year age groups, stratified by sex, up to 1994. These data demonstrate improving survival with successive birth cohorts, and a sex difference in survival, whereby males have improved survival chances. This sex difference narrows with successive birth cohorts, and is particularly marked prior to 1987.

The left censored survival curves apply to prevalent cases alive in 1963. For these people, they again group them into birth cohorts, and infer the proportion of the original birth cohort alive in 1963 at the outset of data collection. This was achieved by collecting UK population birth data for the relevant cohort, and then assuming an incidence of one in 2381 of these births will be CF cases, the incidence calculated in their study. They then applied the same life table approaches to the survivors to calculate survival probabilities going forward in time. The resulting analysis again demonstrates a cohort effect, and also the survivor bias inherent in left censored datasets: the individuals alive in the dataset in 1963 are a group of essentially healthier survivors from the original birth cohort.

The authors go on to apply the age specific mortality rates from the most contemporary birth cohort (1992 to 1994) to the current CF population alive in 2003, and on this basis are able to estimate the current survival for the CF population, estimated to be around 40 years overall, with median survival for men at around 45 years, and that for women at 35 years. They point out, however, that since cohort survival appears to be improving year on year, the estimated median survival for children born since 2000 is likely to be around 50 years. Other key findings from this analysis include the marked reduction in infant mortality in the CF population with successive cohorts, attributed to improved treatment in the first year of life,

particularly of meconium ileus. They also point out that the continued growth of the adult CF population by around 145 patients per annum had national and local implications. This study illustrates some of the technical difficulties inherent in estimating survival in CF populations (Lewis, 1998), and did not allow exploration of the effects of covariates other than sex and birth cohort on survival. The authors further suggest that international comparisons present serious technical and methodological problems, due to different methods of data collection (Dodge et al., 2007).

Recognising the difficulty of undertaking studies of cohort survival across countries due to the lack of standardised data, Fogarty et al compute median age at death for people with CF in 17 countries in Europe, Australia and North America up to 1994 (Fogarty et al., 2000). Trends in median age at death are likely to indicate trends in overall survival. The study shows improving trends across all countries from an international median of eight years in 1974 to 21 years in 1994, but note that that median age of death significantly underestimates median survival in a disease such as CF where cohort survival is improving over time. They also demonstrate marked sex differences, with women having greater chances of dying at a younger age, and apparent differences between countries, with survival chances in the US being the highest, compared to those in Scotland (Odds Ratio (OR), 0.39, 95%CI 0.30 to 0.52 relative to United States the sex-adjusted proportion of people dying from CF at an age above the international median age of death for their year of death). The authors recognize the limitations of the analysis; differences in coding practices between countries; the high sensitivity of the calculation to misclassification of deaths in the first year of life; and small numbers in some countries. Nevertheless they suggest that differences between countries require further investigation, and may relate to socio-economic factors, and access to evidence based treatments (Fogarty et al., 2000).

Finally, in an earlier study of survival from the UK of note, Elborn et al estimated the number of children born each year with CF from 1959 to 1986, and used this together with annual mortality data to generate cohort life tables. They then developed regression models to make forward predictions up to year 2000. These predictions allowed production of life tables for annual cohorts from 1959 to 2000 and hence to estimate the size of the CF population for each year up to 2000. Their

model suggested improving survival in successive birth cohorts, and suggested a linear increase in CF prevalence, predicting a population of around 6000 by the year 2000, with 3400 (57%) aged under 16. The median life expectancy of children with CF born in 1990 was estimated at 40 years, double that of the estimate 20 years beforehand (Elborn et al., 1991).

### ***Nutritional status***

Regarding nutritional issues, Salvatore et al identify 17 studies focussed in this area (Salvatore et al., 2010, Salvatore et al., 2012), with only one from the UK, which explored the relationship between obesity and overweight, and lung function. The study demonstrated a positive association between high BMI, and lung function even at BMI z-score levels of 1 to 2, and the authors argue against calorie restriction in relatively overweight children with CF (Kastner-Cole et al., 2005). The other studies on nutrition have generally confirmed the association of good early nutritional status, with better subsequent lung function, and some have suggested that this is related to reduced risk of death. This evidence suggests that aggressive and early nutritional intervention is particularly important in the first few years of life. The importance of nutrition was first demonstrated in a study comparing outcomes in Toronto and Boston, using the Canadian and US registries, which showed improved growth and survival in the Toronto clinic population, attributed to more aggressive nutritional therapy (Corey et al., 1988), and this was subsequently corroborated in a population level comparison (Lai et al., 1999). This study further demonstrated an equalization of nutritional status between the US and Canadian CF populations, as the US adopted more aggressive nutritional regimens, more in line with the Canadian approach.

A few studies have used multivariate regression approaches to quantify the association between nutritional status, and subsequent clinical status. Zemel et al used a subset of the US registry to show a positive association between nutritional status (weight and height z scores), and longitudinal lung function trajectory (Zemel et al., 2000), results corroborated in Konstan's study, which showed that malnourished children at age three had reduced pulmonary function at age six (Konstan et al., 2003). Furthermore, in one of the larger studies of growth in 19,000 children in the US registry, Beker et al demonstrated a three to five fold increased risk of death in children who were stunted at age five (<5th height centile) (Beker et al., 2001). A study using the German registry further demonstrated that improved

early nutrition was associated with reduced *P. aeruginosa* prevalence, and that malnutrition coupled with *P. aeruginosa* was particularly detrimental to average lung function, and increased lung function decline over a year in a longitudinal analysis (Steinkamp and Wiedemann, 2002).

Taken together with the evidence for newborn screening from registry studies mentioned earlier (Salvatore et al., 2010), the registry studies on nutrition suggest that early nutritional status may play a pivotal role in the chain of events leading from early diagnosis, to later improved lung function, and therefore survival. Furthermore, there are plausible biological mechanisms to support this, since very early life is a critical period for the formation of alveoli in the lungs (Lai et al., 1999). Salvatore et al highlight the importance of further studies that investigate the interaction of early nutritional status, socio-economic status, *P. aeruginosa* acquisition, and nutritional status, especially in settings outside the US, with different health care systems based on universal access (Salvatore et al., 2010).

### ***Lung function***

Between 2001 and 2012, 15 registry studies have focussed on lung function (Salvatore et al., 2011, Salvatore et al., 2012). None of these are from the UK. Here I review the studies of particular relevance to the work in this thesis. The first of these was a large cross-sectional study of 7010 people in the European registry, which showed that a range of factors were associated with a 10 percentage point lower %FEV<sub>1</sub> than expected: low body weight, haemoptysis, pulmonary symptoms at onset, *P. aeruginosa* infection, use of oral steroids, anti-inflammatory drugs, DNase, oxygen requirement and mechanical ventilation (Navarro et al., 2001). Konstan et al undertook the largest subsequent study of risk factors for lung function decline in around 5000 children in the US. The authors prospectively assessed the effect of various risk factors on lung function trajectories over three four year age periods, in children in the US registry. High baseline %FEV<sub>1</sub> and persistent crackles were significant independent risk factors for %FEV<sub>1</sub> decline across all age groups. Female gender, *P. aeruginosa* infection, low weight-for-age, daily sputum production, wheezing, sinusitis, pulmonary exacerbations treated with intravenous antibiotics, abnormal liver test results, and pancreatic insufficiency were also identified as independent risk factors in some age groups (Konstan et al., 2007a).

Vandevanter et al extended this work, to propose a scoring system to identify patients at risk of significant deterioration in lung function, based upon the aforementioned risk factors (Vandevanter et al., 2010). The same authors develop this idea further to suggest that staging of lung disease based on cross sectional measures %FEV<sub>1</sub> needs to be developed further to take into account the more dynamic age related changes in lung function (Konstan et al., 2009). A further study from the US explored factors associated with failure of recovery of lung function following pulmonary exacerbations, including female sex, poor nutrition and persistent *P. aeruginosa* infection (Sanders et al., 2010), and the same group went on to show that increased frequency of pulmonary exacerbations was associated with increased lung function decline in children and adults (Sanders et al., 2011). Liou et al undertook a recent large study of 20,000 people in the ECFS, to quantify annual changes in %FEV<sub>1</sub> in individuals aged 6 to 45 years old. The authors further contrast individual level changes in %FEV<sub>1</sub>, with population level changes. They show that % predicted FEV<sub>1</sub> decreases by one to three points per year for individuals, with maximal decreases in 14 to 15 year olds. Furthermore, there is a large degree of within individual variation. They also show that within individual decline continues in adulthood, but this is not reflected in aggregate measures, which flatten out due to survivor bias in the sample of people over the age of 30 years (Liou et al., 2010).

Two other notable studies have examined the association between environmental factors and lung function in CF. Goss et al demonstrated an association between levels of air pollution and the risk of pulmonary exacerbations, and more rapid decline of lung function in 11,000 people in the US registry (Goss et al., 2004), and Collaco et al demonstrated that exposure to passive smoking was associated with worse lung function in a study of 812 people in the US CFF registry (Collaco et al., 2008). In a more recent study using data from the US, Australia, and New Zealand, Collaco et al further demonstrate an association between warmer temperatures, risk of *P. aeruginosa*, and worse lung function (Collaco et al., 2011).

### **Health inequalities and CF**

The starting point for the studies in this thesis rests on a handful studies of undertaken in the UK and the US, since 1989 (Britton, 1989, Schechter and

Margolis, 1998, Schechter et al., 2001, O'Connor et al., 2003). Though the evidence base has evolved somewhat since starting this thesis (Schechter et al., 2009, Quittner et al., 2010, Stephenson et al., 2010, Barr HL, 2011, Schechter et al., 2011, Taylor-Robinson et al., 2013a), the former remain the key studies on the effect of SES on outcomes in CF. The first observations by John Britton and his group in Nottingham in the UK demonstrated an apparent difference in survival in CF by SES (Britton, 1989). Michael Schechter and colleagues explored this further in the US, first in a single care centre (Schechter and Margolis, 1998), and then in a key publication the following year on a US-wide registry population (Schechter et al., 2001). These studies confirmed a difference in survival on the basis of health insurance status in the US, and also indicated differences in other CF outcomes. Finally, Gerry O'Connor's study, again in the US registry population, demonstrated a social gradient in survival, and also some intermediate CF outcomes (O'Connor et al., 2003). In this section we review each of these key trials in more detail, and other relevant literature, before turning to the gaps in the literature.

Britton's study (Britton, 1989), showing that the chances of age at death at above the median for the general CF population were greater for people in non-manual groups, compared to manual occupations, was motivated by a case-series report from the Royal Brompton Hospital, UK, in 1980 (Penketh et al., 1987). This observational case series reviewed all of the 317 patients attending a specialist centre in London, and found that 56% of patients attending the centre were from social class I and II, using the Registrar General's classification of social class by occupation, compared to an expected 20%. The authors suggested that this may "reflect the type of patients who seek to be referred to a national centre and are able and prepared to travel."

On the basis that these differences in service use by SES may be more widespread in the UK, and that this may result in differences in outcomes between centres, Britton et al analysed death registration data for people with underlying cause of death coded as CF between 1959 and 1986, and for each year calculated the median age at death (Britton, 1989). At this time there was on-going debate about the role of specialist centres in CF, and Britton hypothesised a relationship between the apparent better performance seen in specialist centres and social class effects. They then went on to show that more advantaged individuals were more likely to have a higher age at death in a particular year (ORs for death at age above the median for the year of

death was 1.47, 95% confidence interval (CI) 1.16 to 1.87, non-manual to manual occupations). In the absence of cohort data on which to undertake formal survival analysis to demonstrate premature mortality, this provided initial compelling evidence of differences in survival by SES. Commenting on reasons for the observed differences, whilst accepting the possibility of reverse causation, the authors speculated that this might be due to factors such “lack of resources to permit visits to hospitals or the local doctor or to provide medicines and dietary supplements, or by factors such as increased parental smoking, poor quality or overcrowded housing, and lower levels of education” (Britton, 1989).

Ten years after the Britton study, Michael Schechter and colleagues showed that survival chances were greater in people who never used Medicaid based insurance in the US, compared to those that always used Medicaid, using Medicaid status as a proxy for poverty (Schechter and Margolis, 1998, Schechter et al., 2001). This association was first documented in a care centre based sample of 281 patients aged under 21 years in North Carolina (Schechter and Margolis, 1998), and then followed up in a large-scale study using the US CF registry population (Schechter et al., 2001). In the US, individuals who cannot afford private health insurance are eligible for Medicaid, and the association of Medicaid use with poverty was utilised in this study, since no other indicators of SES were available. The study had two components – a longitudinal survival analysis, and a cross-sectional analysis of other outcomes.

In the longitudinal survival analysis of 20,390 people aged under 20 in the registry from 1986 to 1994, Schechter et al found that people on Medicaid, compared to those never on Medicaid, had a hazard ratio for death of 3.70 (95% CI 3.06 to 4.46) after adjusting for age, race, and genotype, which are known to influence survival, but were not considered to be in the causal path linking SES to survival. After adjustment for level of %FEV<sub>1</sub> at entry to the analysis, however, the effect of Medicaid on SES became non-significant. In the cross-sectional component of the analysis all patients in the registry in 2000 were analysed. Schechter reports significantly lower %FEV<sub>1</sub> at age five in the Medicaid patients (9.2 percentage points, 95% CI 7.1 to 11.4), and this gap increased slightly in an age dependent manner up to age 20.

Furthermore, there was a greater chance of being under the 5<sup>th</sup> centile for weight or height in the Medicaid group. The authors also examined reported use of in-hospital treatments, and found that children using Medicaid, compared to those who never used Medicaid, were more likely to be treated for pulmonary exacerbations with IV antibiotics (OR 1.58, 95% CI 1.27 to 1.96), and were also more likely to spend time in hospital. There was no difference in the age of diagnosis of children, or number of clinic visits by Medicaid status. Whilst acknowledging some of the limitations of using Medicaid as an indicator of low SES – one issue being that sicker patients are more likely to be eligible for Medicaid, and thus there is the possibility of reverse causation – the authors speculate that a range of factors that cluster with poverty may explain the associations: poor nutrition or stress influencing immune function; exposure to outdoor or indoor air pollution, including cigarette smoke; earlier exposure to respiratory viral infections, respiratory syncytial virus; stress effects on family function influencing adherence to medications; difficulties in accessing primary care. Otherwise, the authors suggest that decreased access to specialist care was not playing a major role in generating adverse outcomes.

Gerry O'Connor et al extended the work of Schechter's team by demonstrating a social gradient in the relationship between area-based income in the US, and mortality rates (O'Connor et al., 2003). Studying essentially the same population as in Schechter et al's analysis a few years later (23,817 white patients under 18 years of age between 1991 and 2000 in the CF Foundation Patient Registry), O'Connor et al demonstrated a graded relationship between area based income and risk of death, with a relative risk of 1.44 (95% CI 1.20 to 1.73, adjusted for sex, age at diagnosis, and mode of presentation of CF) comparing individuals in the lowest area-linked income bracket (<\$20 000), compared with the highest (>\$50 000). The study also longitudinally modelled %FEV<sub>1</sub>, and weight centile, and demonstrated a monotonic gradient in these outcomes by income category. Thus children in the lowest income category had a lower %FEV<sub>1</sub> by 5.5% points at age 6 (P <0.001, adjusted for the same covariates as above), and this gap remained fairly constant up the age of 18. A similar pattern was observed with weight centile, with a gap of 7.3 percentage points (P < .001). O'Connor's study also explored differences in the use of treatments recorded in the CF registry, and showed that children from more deprived areas, as measured by area-based income, tended to have increased use of nutritional

supplements (50.7% in the lowest income category, compared to 33.9% in the highest), and were more likely to be screened for CF related diabetes. However, there was no significant trend between income categories and the use of DNase, or inhaled antibiotics (aerosolized TOBI®). The authors point out the limitations of the ecological (area-based) measure of SES used in their analysis, and suggest that the associations observed may be due to poor adherence to medications, or local environmental conditions.

In summary, at the outset of the work in this thesis, one study in the UK had demonstrated a difference in survival for people with CF, on the basis of occupational social class. Two US studies corroborated this finding, demonstrating worse survival in the US population on the basis of both Medicaid status, and area-based income. Furthermore, these studies demonstrated inequalities in other important CF outcomes, such as nutritional status, and lung function in the US, and these differences were socially graded in the O'Connor study. The evidence from the US did not suggest that poor access to specialist care was a major factor in the pathway to adverse outcomes.

## Gaps in the literature

The review of the literature has identified a number of gaps in knowledge of the effects of SES on outcomes in CF.

1. Current studies have focussed on the US population. Two studies in the UK have explored the effects of SES on risk of death, but otherwise there have been no population level studies in the UK, or other European countries, exploring the effect of SES on longitudinal outcomes in CF. Studies should be undertaken in countries with contrasting social and health care policy contexts to those found in the US.
2. Population level studies in the UK characterising the age-related changes in key CF outcomes such as lung function, growth, and risk of *Pseudomonas* colonisation are lacking. Developing an understanding of these age-related changes is a pre-requisite to assessing the longitudinal effects of SES on these outcomes.
3. Previous studies of differential effects by SES in the US or elsewhere have almost exclusively focussed on the paediatric population, and little is known about the effects of SES on outcomes in the adult CF population after the age of 20 years. As people are living longer with CF, understanding the effects of SES on the adult population is increasingly important.
4. No previous studies have used a validated, small-area based measure of social deprivation. The studies using ecological measures in the US are limited by the large number of individuals linked to each zip code, relative to measures available in the UK, such as the Index of Multiple Deprivation (IMD).
5. Given the increasing availability of high quality longitudinal CF registries, with repeated measures on individuals over time, modern longitudinal data analysis techniques appear to be under-utilised. Cross-sectional analyses have been favoured in some studies to characterise age related changes in CF

outcomes by SES, and this is particularly the case for use of CF related treatments, which have not been characterised longitudinally.

6. Previous studies have not been informed by a theoretical perspective, or model for understanding pathways to health inequalities. Perhaps as a result of this, current studies in CF have focussed on health outcomes, and how these differ by SES. There is a knowledge gap around the differential social consequences of CF, by SES, especially with regard to the employment consequences.
7. Little is known about the intermediate factors that may mediate any effects of SES on CF outcomes. Health care, as a mediator, has been studied in the US, but not elsewhere. There are significant gaps regarding behavioural, psychosocial and environmental risk factors that may mediate any SES related effects on CF outcomes.
8. Little is known about the effect of newborn screening for CF in the UK, and whether this may have any differential effects by SES.

This thesis seeks to address some of these gaps in the literature using CF registry data from the UK and Denmark.

## **Summary**

Conceptually, CF is an interesting case for the study of pathways to inequalities in health: as a disease of autosomal recessive inheritance where carriers are unaffected, a socio-economic bias in disease prevalence is not expected (Schechter et al., 2001, O'Connor et al., 2003), but there is potential for the development of a social gradient in health care use, disease outcome and consequences.

The studies described here set out to make an important contribution to this developing agenda around health inequalities. They investigate the mechanisms by which inequalities in health and social outcomes are manifested in CF. The distribution of CF is not socially determined, and any differential outcomes are likely to be the result of the complex interaction of genetic, socio-environmental and health

care factors during childhood and over the life course. The studies aim to improve our understanding of how influences in early life, and the health system in the UK mitigate or contribute to differential outcomes in health and employment outcomes in CF. The studies also set out to generate an in-depth longitudinal description of the health, health care and social outcomes for people with CF, thus contributing to our understanding of the clinical and social epidemiology of CF. The UK study in this thesis scrutinises the degree to which equity of service delivery is achieved for key elements of CF care in the NHS, whereas the Danish study offers the opportunity to analyse a unique longitudinal dataset. In addition to developing methodological approaches, and building research capacity around health inequalities, the findings from this thesis are likely to be relevant to other conditions, and will inform policies to improve patient care more generally within the NHS.

## Chapter 3: Methods

I begin by describing the data sources for the studies in this thesis: the UK CF registry, the sources of social deprivation data in the UK, and the Danish CF registry. I will outline the criteria for extracting the analysis datasets in the UK and Danish studies, along with a description of the size and shape of the final longitudinal datasets. I then move on to a general discussion of longitudinal data analysis techniques, before focussing on the specific methods used in each of the studies.

### UK CF registry

The UK CF Registry is the data source for Studies 1 (clinical and treatment outcomes) and 2 (employment outcomes) in this thesis. The Registry is supported and co-ordinated by the UK CF Trust (Adler et al., 2008, CF Trust, 2013a), and records information about the health and treatment of patients from diagnosis onwards. Over 50 British CF specialist centres routinely collect data in a standardized fashion. Patients attending the British centres are seen in the outpatient clinic for a comprehensive annual review, including evaluation of clinical status, pulmonary function, microbiology of lower respiratory tract secretions, and use of major CF related therapies. The Registry is estimated to include nearly all people with CF in the UK population (Mehta et al., 2004) and is therefore ideally suited to the study of outcomes and treatments across the whole socio-economic spectrum in the UK society.

The registry data have been used previously in a number of epidemiological studies in CF (Adler et al., 2008, Chamnan et al., 2010), but have not been analysed longitudinally before. The data source utilised in this study contains data collected between 1996 and 2010, and has been through rigorous quality control by data managers at the CF Trust, and external consultants at Imperial College, London, who prepare the annual review reports (CF Trust, 2013a). This includes regular monitoring visits to the CF centres to ensure that data entry staff are given training and support; checking of data entries to ensure accuracy; random review of sets of patients' notes; screening for removal of duplicates, and tracking of patient transition from paediatric to adult centres (CF Trust, 2013b). Deaths were verified by checking

with ONS. Data were collected from all UK centres from 1999 onwards, and in 2000, the dataset was estimated to contain biographical information on over 92% of the estimated UK CF population (Mehta et al., 2004), and registrations have increased year on year subsequently. Furthermore the CF Trust have written to every paediatrician and adult chest physician in the UK to obtain data on CF patients, and on this basis the estimated coverage is above 99% (CF Trust, 2013b).

The UK registry started as the UK CF Database, which was established at the University of Dundee, Scotland in 1995. Initially data were collected from 56 paediatric and adult CF clinics, using standardised forms, and validated through a system of double data entry, range checking, and error correction (Mehta et al., 2004). Between 2005 and 2007 the data collection system changed from a paper based return system to utilise the online 'PortCF' software used in the US registry. During this transfer there was extensive retrospective data cleaning and checking, undertaken by independent contractors. At this point, full postcode data were collected for each patient, whereas previously this had been sporadically collected. The UK CF Registry and its current software programme, PortCF, are now in its fifth year with the production of five annual reports (CF Trust, 2013a). Data are collected in over 200 fields. The number of patients for whom a 'complete' data set, defined as the data required to produce the range of clinical outcomes presented in the annual reports, was recorded at 82% in 2009, and this has increased year on year (CF Trust, 2013a), with the figure up to 89% for the latest annual report for 2011 (CF Trust, 2013a). The coverage for core variables such as weight and %FEV<sub>1</sub>, used in the analysis in this thesis is higher, and almost all of the people fulfilling the study inclusion criteria had data in these fields (Figure 8). However, for other variables not routinely summarised in the annual report, the data coverage can be poor. For instance, Port CF contains a field asking about parental education level, but this was completed in less than 5% of individuals.

## **Ethics**

NHS research ethics approval (Huntingdon Research Ethics Committee 07/Q0104/2, available on request) has been granted for the collection of data into the UK database. Each patient provided written informed consent for collection of data in the registry, and for use of anonymised data in research. The CF Trust database

committee approved the use of anonymised data in this study, under the terms of the NHS ethics approval (see letter of support from CF Trust in appendix 3).

### **Entry criteria**

I undertook a longitudinal retrospective cohort study of individuals in the UK CF registry under the age of 40 at last follow-up, with at least one outcome measurement of interest and a valid postal code between 1996 and 2010. The age range for inclusion in the analysis varies between the two studies. For instance, the analyses in Study 1 (clinical and treatment outcomes analyses) are stratified into people <18 ('paediatric' analysis), and the adult population ( $\geq 18$  and <40 years). However, %FEV<sub>1</sub> can only be recorded from around age five onwards, so the analysis for %FEV<sub>1</sub> spans the >5 to <18 age range for the paediatric analysis. For the analysis of employment outcomes I selected ages >20, on the basis that most individuals will have left the education system at this point. Only a small proportion of the data relate to people aged over 40 (5% of the annual reviews). Including these data extends the age range for the analysis up to 78 years of age, and this was excluded for consideration in future analyses.

### **Primary outcomes and covariates**

At annual review, data were collected across the full range of fields in the PortCF system. Appendix 3 contains screenshots from the data entry screens, as viewed by clinicians and members of the MDT as they enter the data. From this dataset, pre-specified primary outcomes and co-variates were selected, informed by the literature review.

### ***Study 1***

The primary clinical outcomes were weight, height, BMI, %FEV<sub>1</sub> and chronic *Pseudomonas* colonisation prevalence. Anthropometric values were converted into standard deviation scores using the UK reference population (Pan and Cole, 2002). Pulmonary function tests were performed according to international recommendations (Miller et al., 2005), measuring FEV<sub>1</sub>, expressed as a percentage of predicted values for sex and height using reference equations from Wang or Hankinson (Wang et al., 1993, Hankinson et al., 1999).

The primary health care outcomes were use of therapies in the previous year (yes or no): IV antibiotics; supplemental nutritional support; DNase; and inhaled antibiotic therapy. Conditional on the use of any IV therapy, I also used the log total number of days on IV therapy as an outcome. Supplemental nutritional support included patients receiving nutritional supplements orally, by nasogastric tube, gastrostomy tube, jejunal tube, or total parenteral nutrition (TPN). Any inhaled antibiotic therapy included Tobramycin solution for inhalation, other inhaled aminoglycoside, Colistin and Promixin.

### ***Study 2***

The primary outcome was any employment in the preceding year (yes or no), which included people recorded as being in either full or part-time employment at annual review.

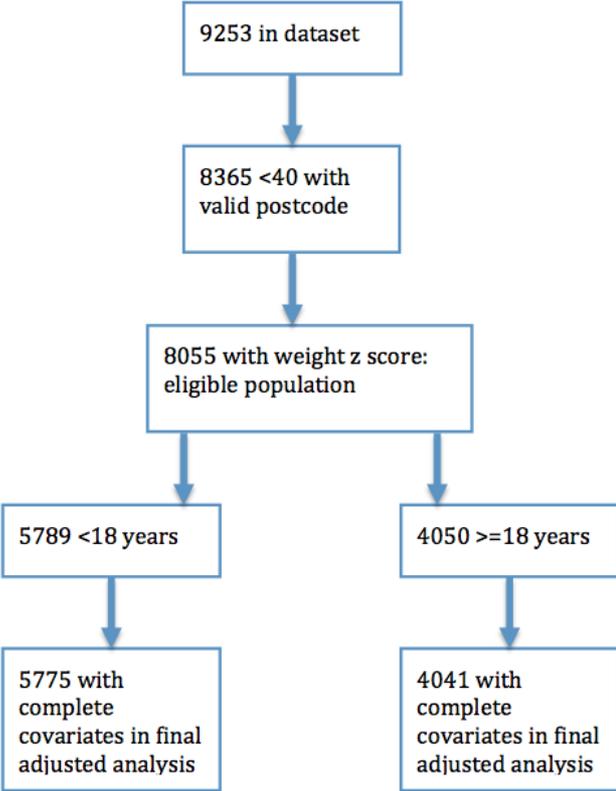
The primary exposure measure of interest in both of the studies was a small-area-based measure of deprivation of area of residence, described in the next section. Postcodes were used to derive IMD scores for the constituent UK countries, facilitated using the online GeoConvert application (GeoConvert, 2011). Each person was allocated to a deprivation score on the basis of the first recorded postcode on entry to the dataset. Other baseline covariates in the analyses included: sex; genotype coded as the number of delta F508 alleles (0, 1 or 2); year of birth; screening status (diagnosis by neonatal screening or otherwise) and ethnicity (Caucasian or otherwise). Time varying covariates included age, presence of CF related diabetes (CFRD) and pancreatic insufficiency (PI), determined by use of pancreatic enzyme supplementation. In the health care use and employment analyses, when it was necessary to make adjustment for disease severity, this was done on the basis of current %FEV<sub>1</sub>, body mass index (BMI) standard deviation (SD) score, and *P. aeruginosa* status where appropriate.

### **Description of analysis dataset**

For the purposes of understanding the structure of the data, we consider the final dataset for the weight analysis here, since this is the most commonly collected outcome in the dataset, collected at 49,337 annual reviews on 8055 patients between 1996 up to 2010 in the UK. The flowchart below (Figure 8) shows the number of

participants included in the analysis, after application of the eligibility criteria. An age based cut off is used to stratify the analysis. People with data straddling 18 years of age can thus contribute to both analyses.

**Figure 8: Flowchart of included participants for weight analysis**



Sixty six percent of individuals had five or more follow up measures (Figure 9), with a mean number of follow-up measures of 6.1 (SD 3.3), and total of 48,425 person-years of follow up. The number of reviews per year increases up to year 2000, and stabilises subsequently, with a slight dip during the process of transfer over to PortCF.

**Figure 9: Data follow-up in the population aged <40 years**

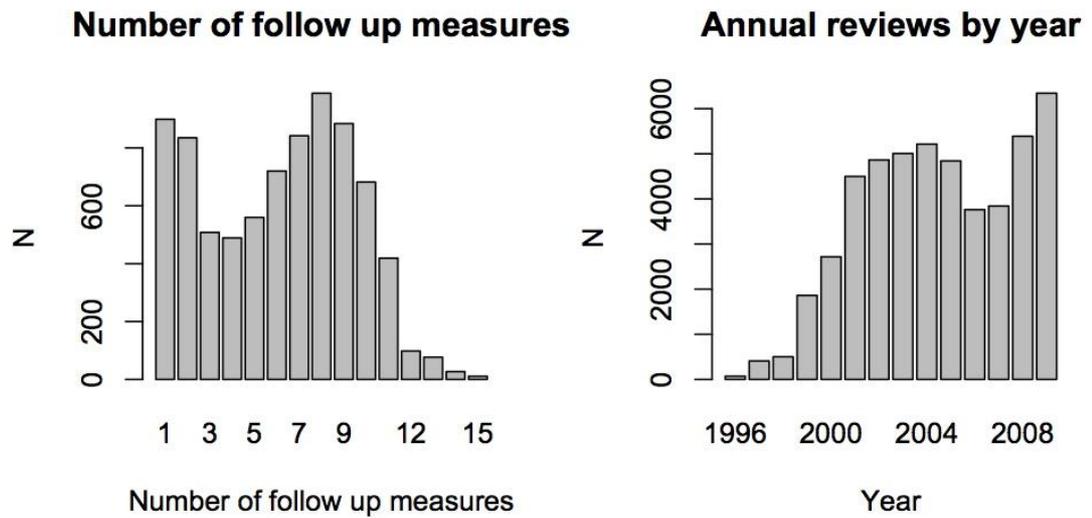
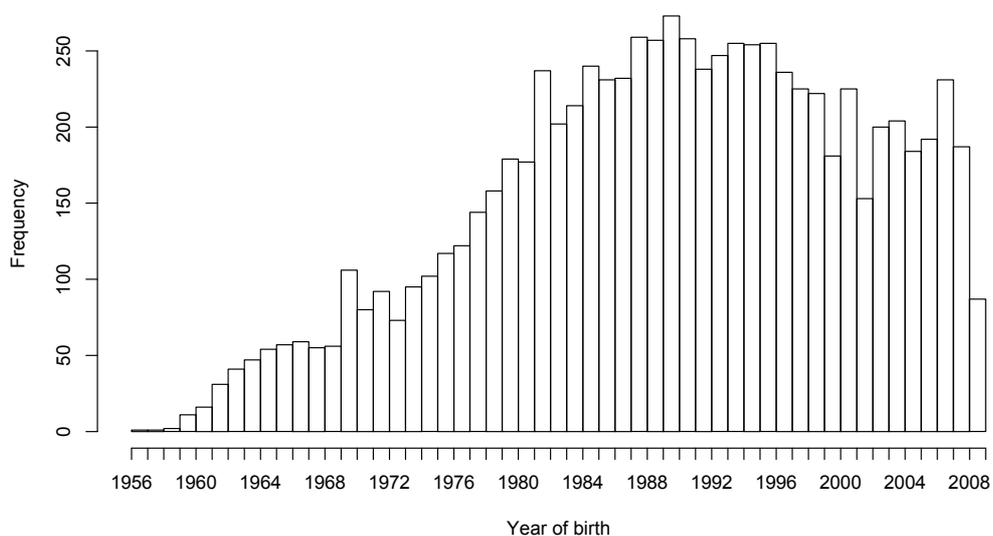


Figure 10 illustrates the distribution of year of birth for people included in the analysis. This illustrates the left censoring in the dataset. The prevalent population captured in the dataset in 1999 onwards, when there was complete UK coverage, represents selected individuals from earlier birth cohorts who have survived to that point. 2066 incident cases are captured subsequently, with birthdates in 1999 and beyond.

**Figure 10: Year of birth for people included in the weight analysis 1996 to <2010**



## Source and extraction of UK SES data

In 2004 all CF centres were asked to collect full postcodes on patients attending for annual review. Completion of the postcode field is mandatory in PortCF and thus coverage of postcode has improved since the switch over to the web based interface. The postcodes were used to link individuals to small area deprivation measures in the UK.

The datasets can be retrieved at the websites listed below, with the exception of the English dataset which is no longer available via the archived website:

- English IMD 2007:  
<http://webarchive.nationalarchives.gov.uk/+http://www.communities.gov.uk/communities/neighbourhoodrenewal/deprivation/deprivation07/> (accessed 26th March 2013). Technical appendix via same site.
- Scottish IMD 2009:  
<http://www.scotland.gov.uk/Topics/Statistics/SIMD/background2simd2009> (accessed 26th March 2013). Technical appendix:  
<http://www.scotland.gov.uk/Topics/Statistics/SIMD/simd2009technical> (accessed 26th March 2013)
- Northern Ireland IMD 2010  
[http://www.nisra.gov.uk/deprivation/nimdm\\_2010.htm](http://www.nisra.gov.uk/deprivation/nimdm_2010.htm) (accessed 26th March 2013). Technical appendix:  
[http://www.nisra.gov.uk/deprivation/archive/Updateof2005Measures/NIMDM\\_2010\\_Report.pdf](http://www.nisra.gov.uk/deprivation/archive/Updateof2005Measures/NIMDM_2010_Report.pdf) (accessed 26th March 2013)
- Welsh IMD 2008 (revised 29/3/2011)  
<https://statswales.wales.gov.uk/Catalogue/Community-Safety-and-Social-Inclusion/Welsh-Index-of-Multiple-Deprivation/WIMD-2008> (accessed 26th March 2013). Technical appendix:  
<http://wales.gov.uk/topics/statistics/publications/publication-archive/wimd2008tech/?lang=en> (accessed 26th March 2013)

The IMD scores for the UK are based on a methodology developed by the Social Disadvantage Research Centre at the University of Oxford, and separate indices have been constructed for England, Northern Ireland, Scotland and Wales (Noble et al., 2006). Indices of multiple deprivation combine economic, social and housing indicators measured at the census into a composite deprivation score for small areas in the UK constituent countries (ONS, 2011a). The indices are based on domains, calculated from census indicators, which are then weighted and combined to create the overall score for each country (Table 2). There were 41,773 of these small areas in the UK, containing on average 1400 people (range 500-3700).

The IMDs in the UK are widely used as measures of SES in epidemiological studies (Semple et al., 2011, Taylor-Robinson et al., 2011, Bergen et al., 2012) and are recommended for tracking health inequalities in UK government statistics (DH, 2012b). The IMD methodology allows much finer resolution than analyses using ZIP codes in the USA, which contain on average 30,000 people (Krieger et al., 2002). I used the postcode first recorded at entry to the dataset to link an individual to an IMD score, in order to generate a fixed measure of area deprivation.

**Table 2: Domains and weights used to generate IMD scores for UK constituent countries**

<b>Domain</b>	<b>England (N= 32482)</b>	<b>Northern Ireland (N=890)</b>	<b>Scotland (N=6505)</b>	<b>Wales (N=1896)</b>
<b>Income</b>	22.5	25	28	23.5
<b>Employment</b>	22.5	25	28	23.5
<b>Health</b>	13.5	15	14	14
<b>Education</b>	13.5	15	14	14
<b>Barriers to housing and services</b>	9.3	-	-	-
<b>Proximity to services</b>	-	10	-	-
<b>Geographic access</b>	-	-	9	10
<b>Housing</b>	-	-	2	5
<b>Living environment</b>	9.3	5	-	-
<b>Physical environment</b>	-	-	-	-
<b>Crime</b>	9.3	5	5	5

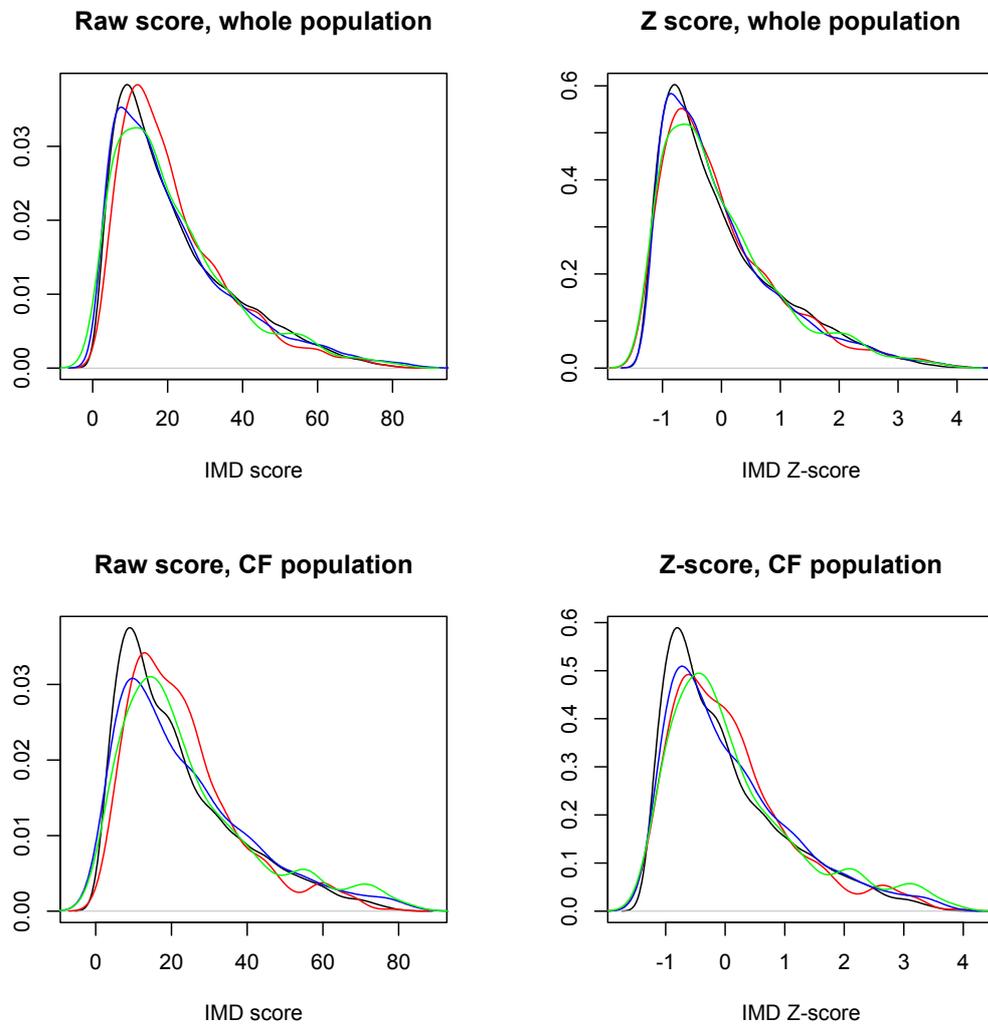
*Adapted from ONS 2011 (ONS, 2011a)*

Although based on the same concept and general methodology, there are differences between countries in the weighting of the domains, the indicators used to generate the domains, the spatial scale at which the indices are calculated and the time points on which they are based (Noble et al., 2008). For instance the number of indicators used to construct each domain varies between countries, with Scotland using 37 indicators, compared to 52 in Northern Ireland. Furthermore, the small areas on which the variables are derived are of similar size in England, Wales and Northern Ireland (approximately 1500 people), but are smaller in Scotland (approximately 750 people) (ONS, 2011a). However, Figure 11 (top left panel) shows that the distribution of deprivation scores in the four countries is very similar.

Since the composition of the deprivation scores for the UK constituent countries differs slightly, and following the advice of Peter Goldblatt, I explored standardising the IMD scores (subtracting the within-country mean, and dividing by the within-country standard deviation), in order to generate more directly comparable measures. The distributions of the raw and standardised scores for the whole population, and the CF population of each country are shown in Figure 11.

**Figure 11: Density plot comparing distributions of deprivation scores**

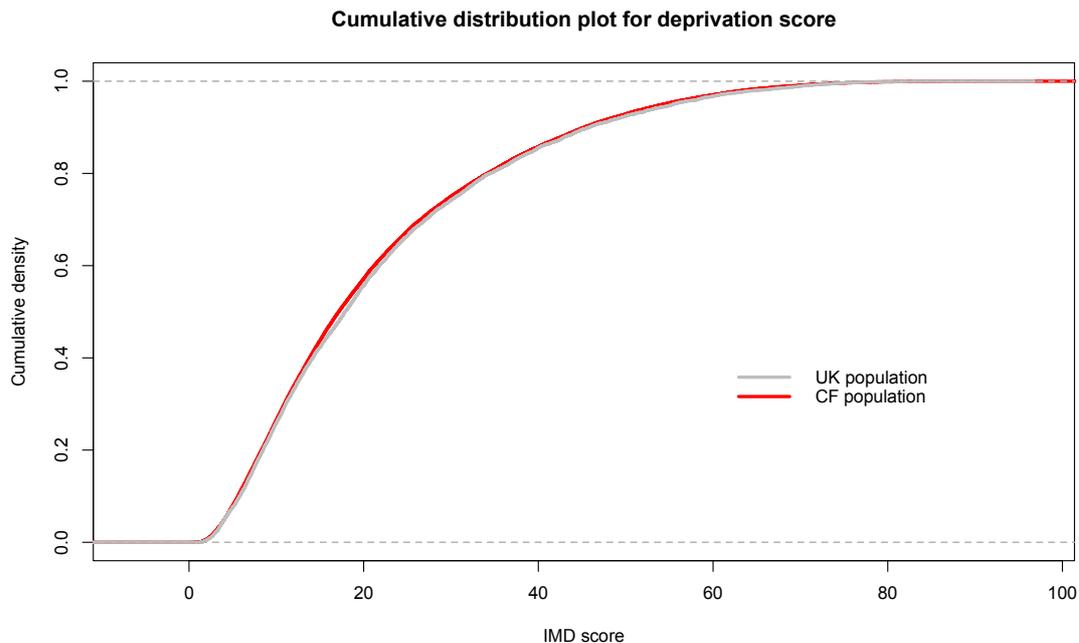
*The top row shows the whole populations (England in black, Scotland in blue, Northern Ireland in green and Wales in red), raw scores and standardised. The bottom row shows the CF populations.*



All of these small areas were then ranked on the basis of the continuous deprivation SD (or Z score), and then divided into fifths, or 'quintiles', providing the following approximate cut-off points for normative deprivation quintiles. When comparing the results using this procedure, to those generated using unstandardized raw deprivation scores, this made little substantive difference, and the distributions were almost identical. Furthermore the distribution of IMD scores was indistinguishable on a cumulative density plot for the UK population, and the UK CF population (Figure 12). For ease of interpretation, raw deprivation scores were used in the final analysis, using quintile cut-off points as follows: <8.31; 8.32 to 13.81; 13.82 to 21.20; 21.21

to 34.11, >34.11. For instance, the small area (lower super output area) linked to the postcode where I live has a deprivation score of 16.1, placing me in the middle deprivation quintile, in terms of deprivation of area of residence.

**Figure 12: Cumulative density plot comparing deprivation score distribution for the whole UK population versus the CF population**



In the analysis, although IMD is measured on a continuous scale, for descriptive summaries I have followed the common practice of grouping IMD into quintiles. However, reducing IMD to a categorical variable loses information, and also leads to models that are difficult to interpret, especially when this five-level categorical variable interacts with non-linear time effects. Thus for the analyses in the studies of clinical outcomes, and treatments, I retained IMD as a continuous variable, and the fitted beta coefficients for IMD score were then used to summarise the effect of deprivation by comparing a person in the mid-point of the most deprived quintile to one in the mid-point of the least deprived quintile.

### **Danish dataset**

CF care in Denmark is delivered through two centres in Copenhagen and Aarhus. The Danish CF Patient Registry is one of the longest running CF registers. It is estimated that coverage of people with CF resident in Denmark is almost complete

from 1990 onwards, when CF care was centralised. This coverage and the unparalleled frequency of measurement (monthly) make this a unique dataset for epidemiological research (Taylor-Robinson et al., 2012a).

The registry started in Copenhagen in 1989 and initially captured all patients alive in 1989 under the care of the Copenhagen centre (around 300 patients), with individual lung function measures recorded monthly, some going back as far as the late 1960s. From 2001 the Copenhagen registry was merged with data from the Aarhus centre (around 150 patients), and thus achieved a full record of the whole of the Danish CF population (personal communication, Dr Tania Pressler). The register is administered in the Rigshospitalet in Copenhagen (Dr Tania Pressler), and Aarhus University Hospital Skejby (Professor Oluf Schiøtz and Hanne Verbert Olesen) (Figure 13).

**Figure 13: Map of Denmark showing location of CF centres**



Source: <http://www.mapsopensource.com/denmark-map.html>

At the outset my intention was to analyse data from the Danish CF register, using similar approaches to those used to analyse the UK data, but utilising the individual level SES data available in Denmark, particularly measures of parental education. Data linkage to social registers in Denmark, facilitated through Statistics Denmark on the basis of a unique identifier for everyone in the population, allows access to richer individual level socio-economic data than is available to epidemiologists in the

UK, on both patients and their parents. These linked data extend retrospectively to 1981 and contains information on employment, income, healthcare expenditure, hospitalisations, and receipt of welfare benefits (Rasmussen et al., 2006, Carlsen et al., 2007, Rasmussen et al., 2007, Carlsen et al., 2008a, Carlsen et al., 2008b). However, due to the very frequent monthly follow-up of CF patients in Denmark, it was necessary to develop a new approach to analysing the Danish dataset, and thus the focus of the analysis presented in this thesis is around the methodological challenges of modelling lung function decline in the Danish population. Research on the effect of SES on CF outcomes in Denmark is on-going.

### **Ethics**

The study using the Danish data was approved by the Danish Data inspectorate – Datatilsynet (see appendix 3).

### **Entry criteria**

All patients aged over five contributing %FEV<sub>1</sub> data in the Danish CF database between 1969 and <2011 were eligible. Entry to the registry was contingent upon a diagnosis of CF made upon the basis of either two known CF-causing mutations in the CFTR gene, and/or two positive sweat-tests together with symptoms compatible with the disease. Post-transplant data from patients who had received a lung transplant were excluded.

### **Primary outcomes and covariates**

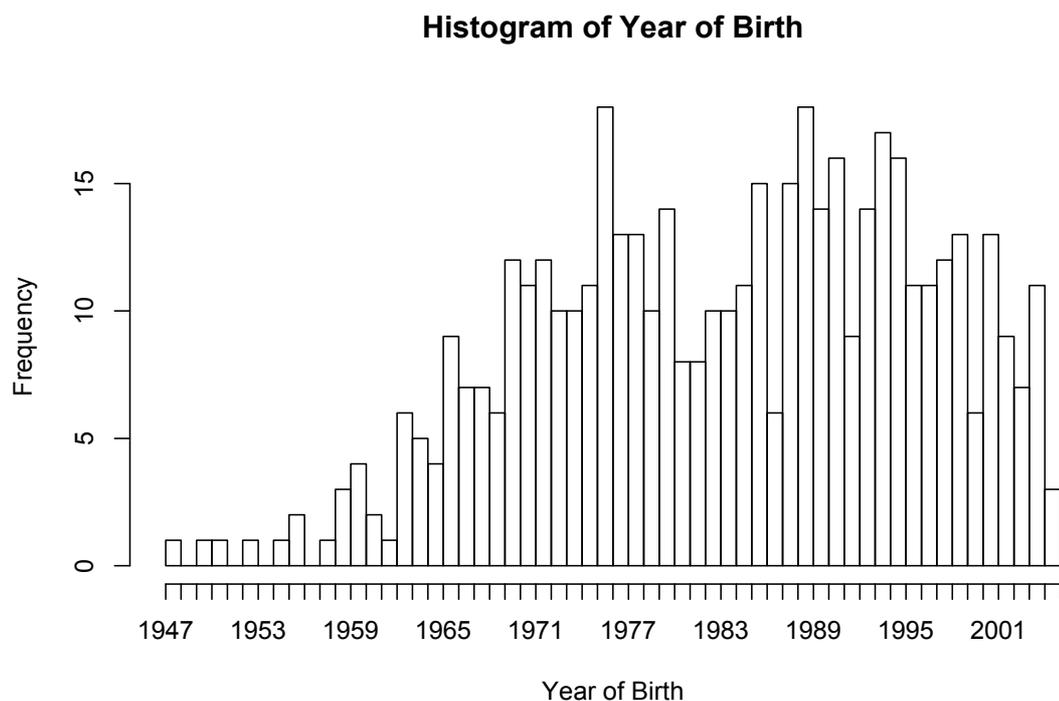
Patients attending the two Danish CF centres (Copenhagen and Aarhus) are seen routinely every month in the outpatient clinic, for evaluation of clinical status, pulmonary function, and microbiology of lower respiratory tract secretions. The primary outcome for the analysis in this thesis is %FEV<sub>1</sub>. Pulmonary function tests were performed according to international recommendations (Miller et al., 2005), measuring forced expiratory volume in one second, expressed as a percentage of predicted values for sex and height using reference equations from Wang or Hankinson (Wang et al., 1993, Hankinson et al., 1999). Covariates in the analysis were: age; sex; genotype coded as the number of delta F508 alleles (0, 1 or 2); onset of chronic *Pseudomonas* infection (coded 0 or 1 as a time-varying covariate); PI

determined on the basis of pancreatic enzyme usage (coded 0 or 1 as a baseline covariate); birth cohort (six 10-year cohorts starting at 1948); and CFRD diagnosed using the WHO criteria (coded 0 or 1 as a time-varying covariate).

### Description of analysis dataset

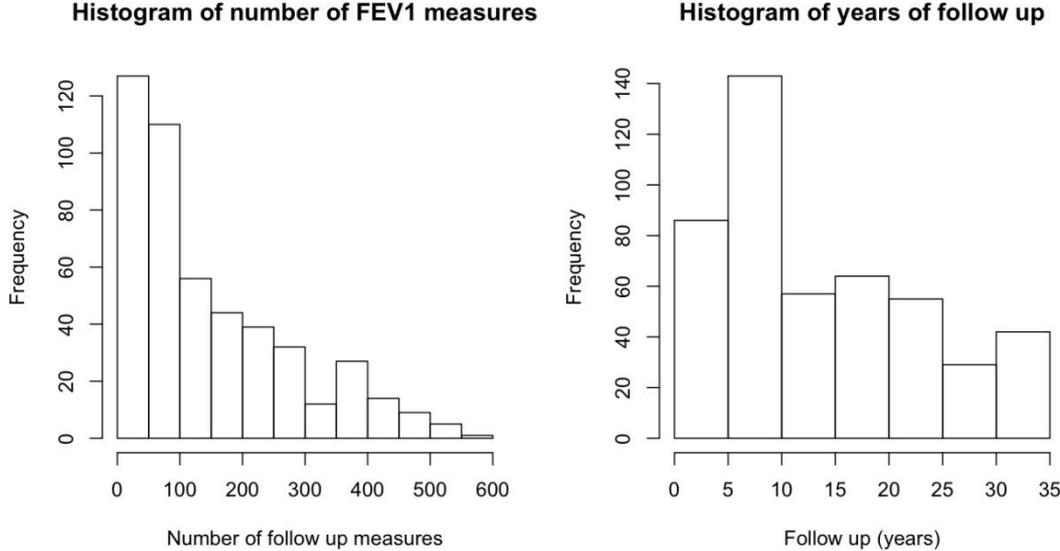
The dataset for the final analysis contains 70,448 lung function measures on 479 patients seen between 1969 and 2010 in Denmark. The median number of %FEV<sub>1</sub> measures per person was 101 (range 2 to 597). Figure 14 is a histogram of year of birth for the people in the dataset. There is evidence of selective recruitment to the dataset in the earlier cohorts.

**Figure 14: Histogram of year of birth for Danish dataset**



The median follow-up period was 10.5 years (range 0.1 to 31.5), with a total of 6500 person-years of follow up. 42 patients were followed up for more than 30 years. Figure 15 shows the frequency and length of follow up for people in the Danish dataset.

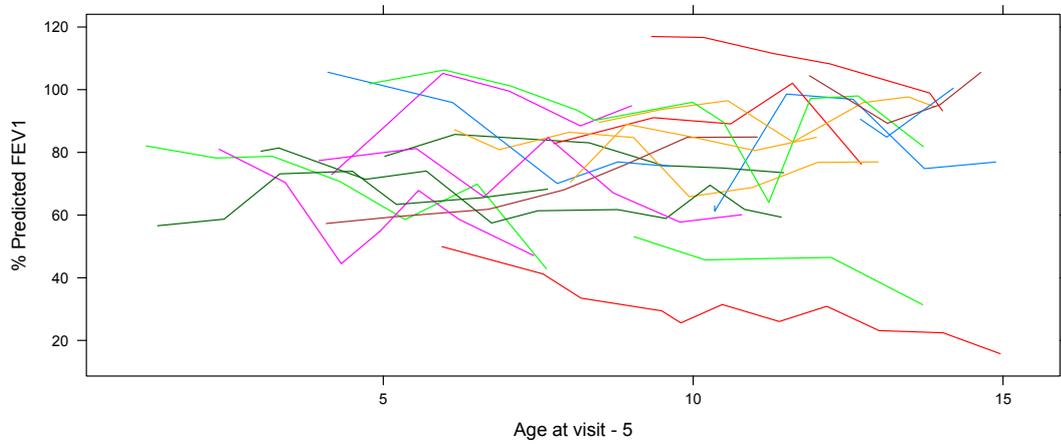
**Figure 15: Histogram of frequency and duration of follow up**



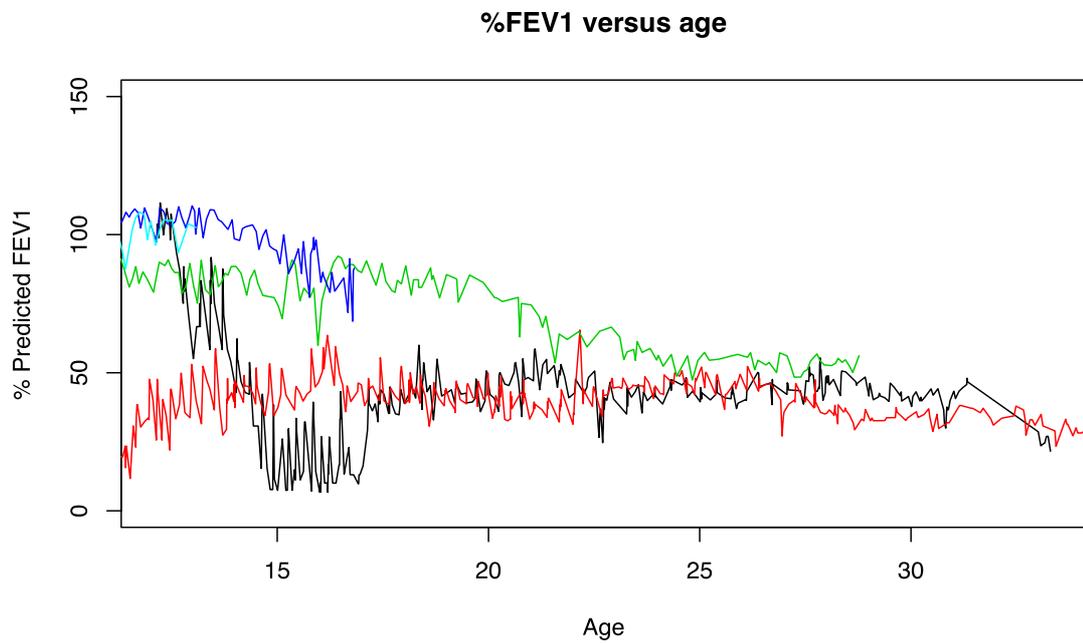
**Contrasting UK and Danish datasets**

The dataset for the UK weight analysis contains 49,337 measures on 8055 people with CF between 1996 and <2010, whereas the dataset for the Danish analysis contains 70,448 measures on 479 people between 1969 and <2011. The UK dataset thus contains short data traces, on many individuals, whereas the Danish dataset contains long data traces on a smaller number of individuals. Without any further analysis, these different data structures suggest that the UK dataset is likely to be better powered to identify cross-sectional differences between sub-populations, since there are more individuals in the analysis. However, the very long data-traces in the Danish dataset make it particularly suited to tracking changes in individuals over time. The spaghetti plots below illustrate randomly selected individual %FEV<sub>1</sub> traces from the UK and Danish datasets (Figure 16, Figure 17).

**Figure 16: Twenty randomly selected %FEV<sub>1</sub> profiles from the UK CF registry for people aged <20**



**Figure 17: Five randomly selected %FEV<sub>1</sub> profiles from the Danish registry**



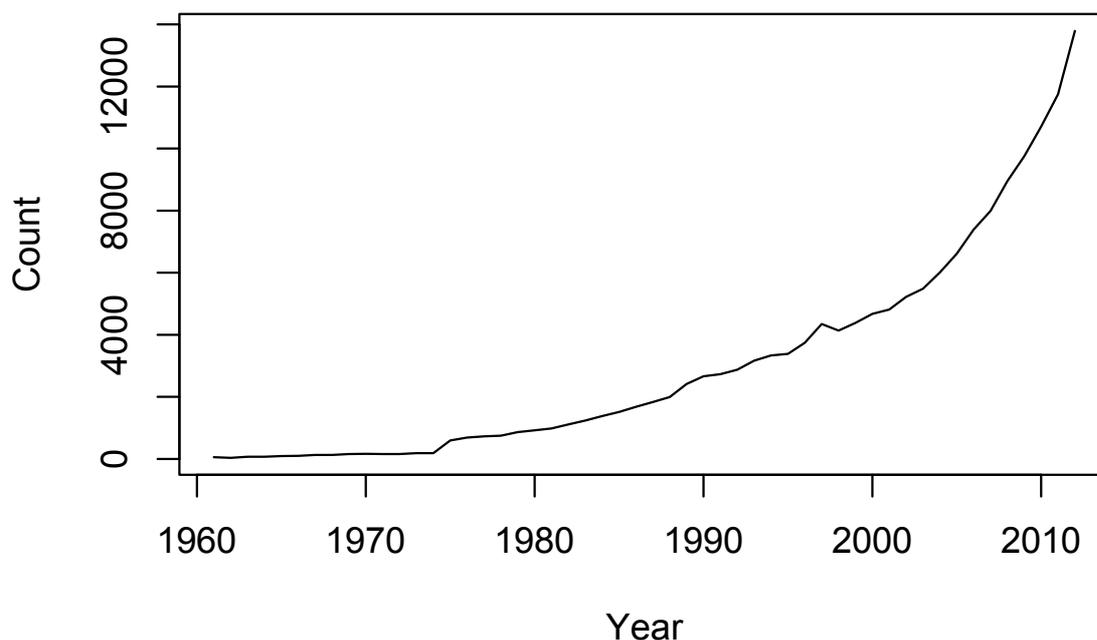
## Statistical methods

In the studies in this thesis I use generalized linear mixed models for the analysis of repeated measurements over time. These methods allow examination of association between important binary and continuous CF outcomes and SES whilst allowing for correlation within patients, trends over time and missing values (Diggle et al., 2002). This section begins with an overview of some of the general features of longitudinal data, and a description of common approaches to analysis, before describing in detail the specific methods used in the studies.

### Longitudinal data

Longitudinal data is ubiquitous in the biomedical sciences, but the statistical tools for analysing such data were limited, prior to the rapid development of methods, and widely available computing power since the 1980s. Despite this, the application of these modern data analysis techniques in the medical sciences, public health and epidemiology has tended to lag behind the statistical innovation, and studies have often used statistical techniques that fail to adequately take account of longitudinal study designs (Fitzmaurice and Ravichandran, 2008). As illustrated in Figure 18, it is only relatively recently that the use of longitudinal studies has really taken off.

**Figure 18: PubMed results containing “longitudinal” by year, 1960-2012**



## **Characteristics of longitudinal data**

A longitudinal study is one in which measurements are collected repeatedly over time on each subject in the study. The *raison-d'être* for such analyses is typically to characterize the changes in the response of interest over time. The subject in the case of the studies in this thesis is an individual with CF.

There are a number of general features of longitudinal data, which add richness to the types of analysis that can be undertaken but also increase complexity. Correlation of measures within individuals is the key feature that sets longitudinal studies aside from cross-sectional approaches. In longitudinal studies, we generally assume independence of subjects, but we cannot assume that repeated measures on the same subject are independent. Since most biological processes are not completely random, knowledge of the value of a response on one occasion provides information about the likely value of the response on a future occasion. For example, how one is feeling at a particular moment, is generally informative as to how one will be feeling a few hours hence, and generally measurements closer in time within an individual are likely to be more similar than those farther apart in time. Without this correlation between values, there can be no forward prediction, but this correlation between individuals needs to be modelled appropriately in order to make correct inferences from the data (Singer and Willett, 2003, Twisk, 2003).

The average profile of change over time for a population can be complex, and individuals may exhibit considerable variability, which can change over time, so that the variance of the response changes over the duration of the study. Coupled with the correlation between individuals, these features violate the fundamental assumptions of independence and homogeneity of variance that are the basis of many cross-sectional analysis approaches, such as multiple regression, or t-tests (Fitzmaurice and Ravichandran, 2008). Furthermore the data themselves may be collected in an 'unbalanced' fashion, at unequally spaced intervals, and contain missing values. This is often the case even if the study design was intended to collect data in a balanced fashion, due to the practical difficulties of real-life follow up of large numbers of individuals over time, since there will almost always be people who miss their scheduled visit or date of observation. In the UK CF registry, for example, follow up is nominally on an annual basis, but the data are unevenly spaced, with different

follow-up times for individuals throughout the year. Missing data, considered in more detail later on, are a ubiquitous problem in longitudinal studies, and there are various approaches to addressing the issues raised by missing data. For example, complete case analysis is a common and simple method for handling incomplete data, whereby individuals with incomplete datasets are excluded from the analysis. However, this approach can not only be inefficient, but may also introduce bias, when the individuals who are excluded are not a random sample from the target population (Altman and Bland, 2007). Some of the earlier methods for analysing longitudinal data, such as multivariate analysis of variance for repeated measures (MANOVA) are restricted to datasets with the same, complete follow-up protocol for each subject. This requirement for fixed number of repeated measurements on all study participants at a set of common time points essentially meant shoe-horning the data to fit the analysis strategy, and making unrealistic assumptions (Weiss, 2005). Modern techniques have now eclipsed these approaches, by dealing with unbalanced follow-up protocols and missing data in a more sophisticated manner, with less stringent assumptions.

### **Advantages of longitudinal data analysis**

In observational longitudinal studies, a measure on an individual at a particular time-point can be influenced by age (time from birth), period effects (the date), and birth cohort effects (characteristics experienced by individuals born at a particular time (Twisk, 2003). Knowing two of these measures defines the third: one's age today and birthdate define the date today. Longitudinal studies allow one to separate these effects: age related processes within subjects (time-trends), from cross-sectional effects between subjects (cohort effects), whereas in cross-sectional analyses these are conflated (Diggle et al., 2002). In the case of CF, the literature review has highlighted the significant cohort effect, whereby the health of people with CF has improved incrementally over time (Dodge et al., 2007, Buzzetti et al., 2009). This process needs to be separated from the age-related trend in the opposite direction towards decreasing health over time; only longitudinal designs allow this.

As argued in the previous chapter, this thesis aims to explore the consequences of SES on outcomes in CF, which is a disease that we hypothesise does not discriminate by SES in terms of incidence. In order to study these effects from a life course

perspective, considering the evolution of outcomes within groups of individuals over time, then longitudinal data are required. A longitudinal approach is essential to establish the direction of causation, and temporal ordering of influences over the life-course (Kuh et al., 2004). Longitudinal data are also required to unpick mechanisms of social selection, versus those of health selection, described in the previous chapter. A major limitation of cross-sectional analyses, therefore, is that they do not allow conclusions about the direction of causality, because data are only collected at one point in time. In a cross-sectional study of lung function in adults with CF, for instance, if one observes a relationship with SES, it is not possible to determine if low SES has resulted in poor lung function, or vice versa. Schechter et al undertook a cross sectional analysis of the age related changes in lung function in population at different ages (Schechter et al., 2001). In this analysis, lung function measures at different time periods are pooled together, conflating age related changes, cohort effects, and dropout effects. Therefore this does not directly address the effect of ageing on a particular individual's lung function, which requires the collection of data over time (Nakai and Ke, 2009).

Longitudinal studies allow one to assess the influence of covariates on both average responses, and rates of change of response over time in subjects, and open up the possibility of prediction. In the studies here we are particularly interested in exploring average differences in outcomes and time-trends between groups of individuals on the basis of their SES. Longitudinal studies allow one to separate these trends over time within individuals from cross-sectional differences among individuals. In this respect each person in the study can be thought of as serving as her or his own control, and the influence of hidden or unobserved factors on the same person is cancelled out over time, allowing for quantification of the effect of ageing on outcomes of interest.

### **Simple methods of longitudinal data analysis**

Approaches such as analysis of variance (ANOVA), ANOVA for repeated measurements, and derived variables methods are well described, and are technically simple to apply, in that no advanced computation is required (Diggle et al., 2002). They are mentioned here only briefly, since they are not used in this thesis.

Considering a continuous outcome, time-by-time ANOVA, involves undertaking a separate analysis at each particular time point, and thus assumes a common follow-up protocol for each individual in the study. Similarly, if one was interested in comparing the mean response over time between two groups, one could undertake repeated t-tests at successive common time points. These approaches are simple, but lead to multiple inferences over time that are not independent, and thus cannot be easily combined (Diggle et al., 2002). Furthermore they fail to make use of the efficiency gains that are achieved by explicitly modelling the covariance structure. In repeated measures ANOVA, the population mean over time is modelled as a function of group, time, and group-by-time interaction effects, with time treated as a categorical factor in the analysis. Again this requires a complete, and common follow-up protocol for each individual, which is a major drawback. It assumes a constant correlation between repeated measures on an individual, regardless of the time-interval between observations. These approaches are not suited to assessing the effects of multiple covariates on outcomes, nor can they provide information about individual level change over time (Twisk, 2003, Fitzmaurice and Ravichandran, 2008).

Another simple way of analysing a longitudinal dataset is to derive summary variables for each individual, and thus reduce the problem to a cross-sectional one. For example, a continuous outcome trajectory over time for an individual might be modelled as a straight line in the simplest case, and summarised as an intercept value, and a slope.

These derived variables can then be treated as separate outcome measures in a cross-sectional regression analysis in order to assess the effect of baseline covariates. This approach assumes that an individual data profile can be summarised in a way that addresses the hypothesis in question, which is not always the case. Furthermore, it becomes invalid if there is missing data, or different numbers of follow-up measures, because the assumption of common variance for all observations is no longer satisfied. Methods have been proposed to weight the analysis, with data from each subject used to determine the weights (Matthews, 1993), but others have cautioned against this because it ignores the correlation structure of the underlying data (Diggle et al., 2002).

These ‘simple’ approaches to longitudinal data analysis (LDA) only estimate and compare the group means. They are not informative about individual level change. They have now been made somewhat redundant by the modern LDA approaches described below which can be readily implemented in software packages, and are not limited to balanced data, with largely complete follow-up.

### **Modern longitudinal data analysis methods**

These methods are all forms of a generalized linear model (GLM) for longitudinal data and can be categorized into three groups: marginal, random effects, and transition models (Diggle et al., 2002). These approaches can be applied to linear, binary and ordinal outcome measures, by using the appropriate link function – the function that links predictors to outcomes. For example a binary outcome can be transformed using the logit link, and then modelled as a GLM (Weiss, 2005).

To account for the features of longitudinal data described above, these models have two main components; a model that describes the effects of covariates on the mean response over time coupled with a model for the covariance among repeated measures (Fitzmaurice and Ravichandran, 2008). Which modelling option one chooses is partly determined by the hypothesis in question. Often the main focus of interest is in understanding factors that influence the mean response over time, but the inferences made here are dependent on choosing a robust model to take into account the covariance structure. By contrast, understanding the covariance structure itself may be the target of scientific interest, as in the study of lung function decline using the Danish dataset in this thesis (Taylor-Robinson et al., 2012a). In this study I develop a parametric model for the covariance structure of %FEV<sub>1</sub> measures over time, which provides a description of how lung function changes within an individual with CF over time, and quantifies the predictive value of a baseline lung function measure for subsequent measures over time.

Marginal, or population-averaged approaches to LDA are suitable when the hypothesis of interest relates to differences between groups or sub-populations that share common variables. In these models, the marginal expectation (sub-population average) is the primary focus, and the correlation is treated as a nuisance factor. Generalized estimating equations (GEEs), introduced by Liang and Zeger (Liang and

Zeger, 1986), represent a popular marginal approach to LDA, where the basic premise is that one is primarily interested in the regression parameter for the mean response, and not the variance-covariance matrix of the repeated measures. In GEE a ‘working’ correlation structure is assumed, which ideally should be consistent with the structure of the data, since an incorrect choice of correlation structure reduces efficiency (Nakai and Ke, 2009). However, with large samples, this becomes less important, and the GEE approach becomes robust to mis-specification of the working correlation structure (Diggle et al., 2002). In practice, this means one can focus on the model for the marginal mean, and then try different correlation structures whilst checking that the parameter estimates and standard errors remain broadly consistent.

Mixed- or random-effects models are used in the studies in this thesis, and are described in further detail with full statistical notation in a subsequent section. The primary contrast with marginal approaches is that random-effects models allow inference at the individual level (Diggle et al., 2002). These models focus on the regression relationship between outcomes and covariates for an individual, by incorporating two distinct components: A fixed-effect component, which describes covariate effects that are assumed to be identical for each subject, and thus common across the population; and a random-effects component, that extends the model to multiple individuals, by allowing each individual to vary from the population average in a prescribed manner. This heterogeneity between individuals reflects natural variability due to unmeasured factors that are not captured in the fixed-effects. The fixed-effects in the model are estimated as discrete parameters, whereas the random-effects are drawn from a probability distribution, the characteristics of which are estimated in the model. Measurements on the same individual share the random-effects for that individual. The model also contains an error term, and together with the assumptions made about the random-effects this determines the correlation structure in the model (Diggle et al., 2002, Singer and Willett, 2003, Twisk, 2003, Weiss, 2005).

Transition or Markov models appear to be less commonly used in epidemiology. In brief, these models incorporate lagged effects, whereby correlation exists because past outcome values explicitly influence the present observation. Thus the covariates and previous outcomes are included in the model as predictors for the current

outcome. These models are suitable when you believe that the distribution of the current measurement depends directly on earlier measurements (Diggle et al., 2002).

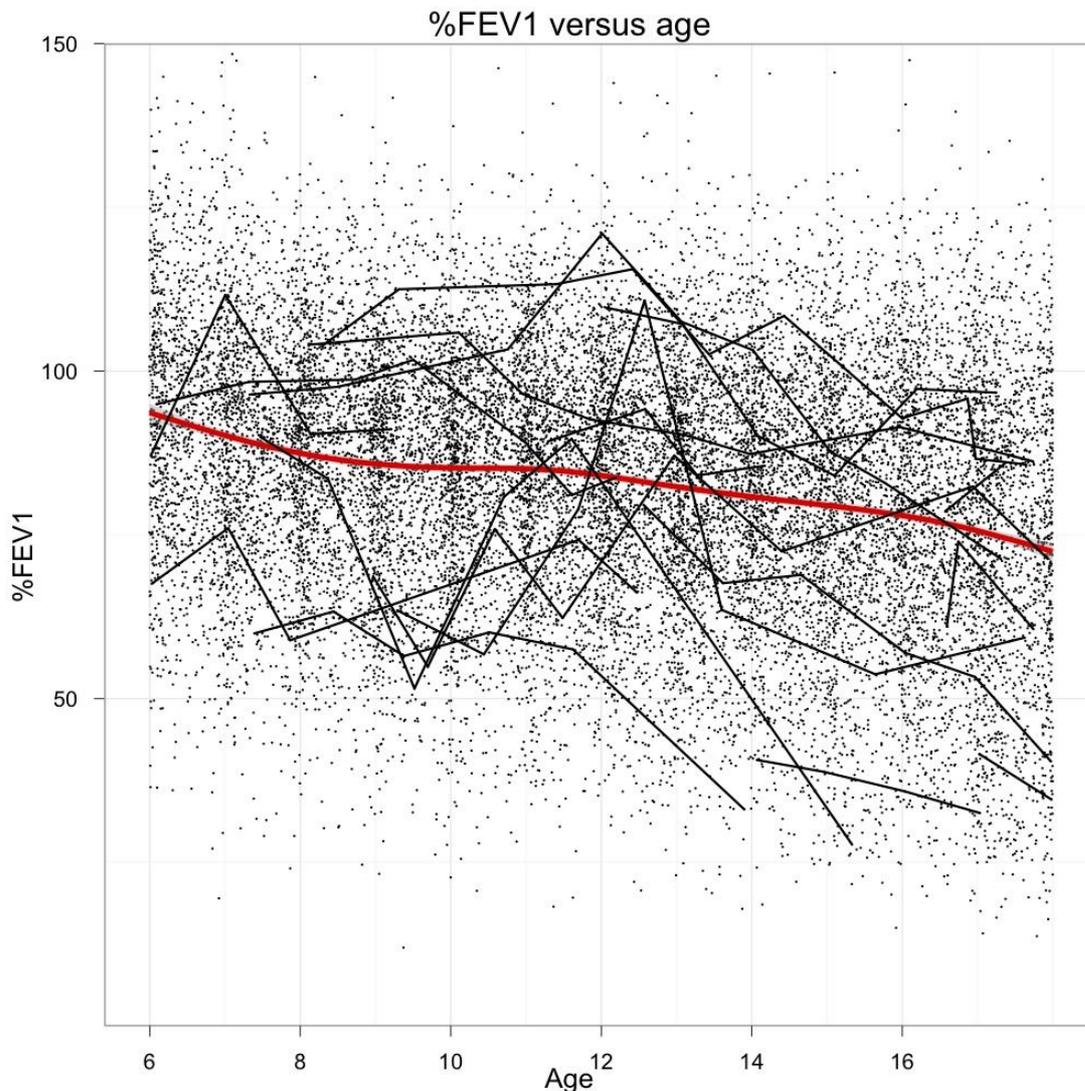
### **Exploratory methods**

Longitudinal data analysis, like other statistical analyses, begins with exploration of the data before moving on to confirmatory analysis. A key step is to visualise patterns in the data that are relevant to the hypotheses in question.

For the analysis of continuous outcomes (e.g. weight, height, BMI, %FEV<sub>1</sub>, IV days in this thesis), the first step is to plot a basic scatterplot of the data for all individuals against time. We can then add lines to connect repeated measurements on individuals, to generate a so-called spaghetti plot (Figure 19). This becomes very overcrowded in a dataset with over 8000 individuals, so a common approach is to randomly select a subset of the data, thus thinning out an otherwise very large dataset. These plots can be further augmented by the addition of a smoothed cross-sectional mean, which can be generated using a kernel smoother, more details of which are outlined below (Diggle et al., 2002). All of these features can be seen in Figure 19, which provides an initial exploratory analysis of lung function decline in the paediatric CF population in the UK.

**Figure 19: Spaghetti plot for %FEV<sub>1</sub> versus age illustrating the longitudinal nature of the data.**

*Each dot represents a %FEV<sub>1</sub> measure on a person in the dataset. The smoothed population average is shown in the red, and randomly selected individual trajectories are in black.*



This demonstrates the unbalanced nature of follow up for individuals in the dataset, with patients entering the analysis at different ages, and being followed up for different time periods. Overall one can visualise the age related decline in lung function, seen in both the scatterplot element of the figure, and then captured by the smoother mean in red. Plots of this kind thus allow one to determine the provisional model mean trajectory, which in the case of %FEV<sub>1</sub> in the UK paediatric analysis is modelled as a straight line, whereas more complicated mean responses are necessary

for other outcomes, such as piecewise or broken-stick functions for weight and BMI. However, individuals deviate markedly from the smoothed mean: There are clear cross-sectional differences between individuals; and differences in the rate of decline of %FEV<sub>1</sub> over time.

Following exploratory analysis, I move on to fitting mixed-effects models to the data, the specific details of which are described in the next section.

## Random-effects models

Repeated measures on individuals are correlated, and this must be accommodated to obtain valid inferences. To analyse the continuous-valued outcomes in the UK analysis (weight, height and FEV<sub>1</sub>) I used a linear mixed model (Diggle et al., 2002).

Specifically, denoting by  $Y_{ij}$  the  $j$ th repeated measurement on the  $i$ th individual and  $t_{ij}$  the age at the time of measurement, I assumed that:

$$Y_{ij} = \mu_{ij} + U_i + V_i t_{ij} + Z_{ij} \quad (1)$$

$\mu_{ij}$  are the expectations of the  $Y_{ij}$  and are described by a multiple linear regression model, so the mean response is specified as a linear combination of explanatory variables, hence

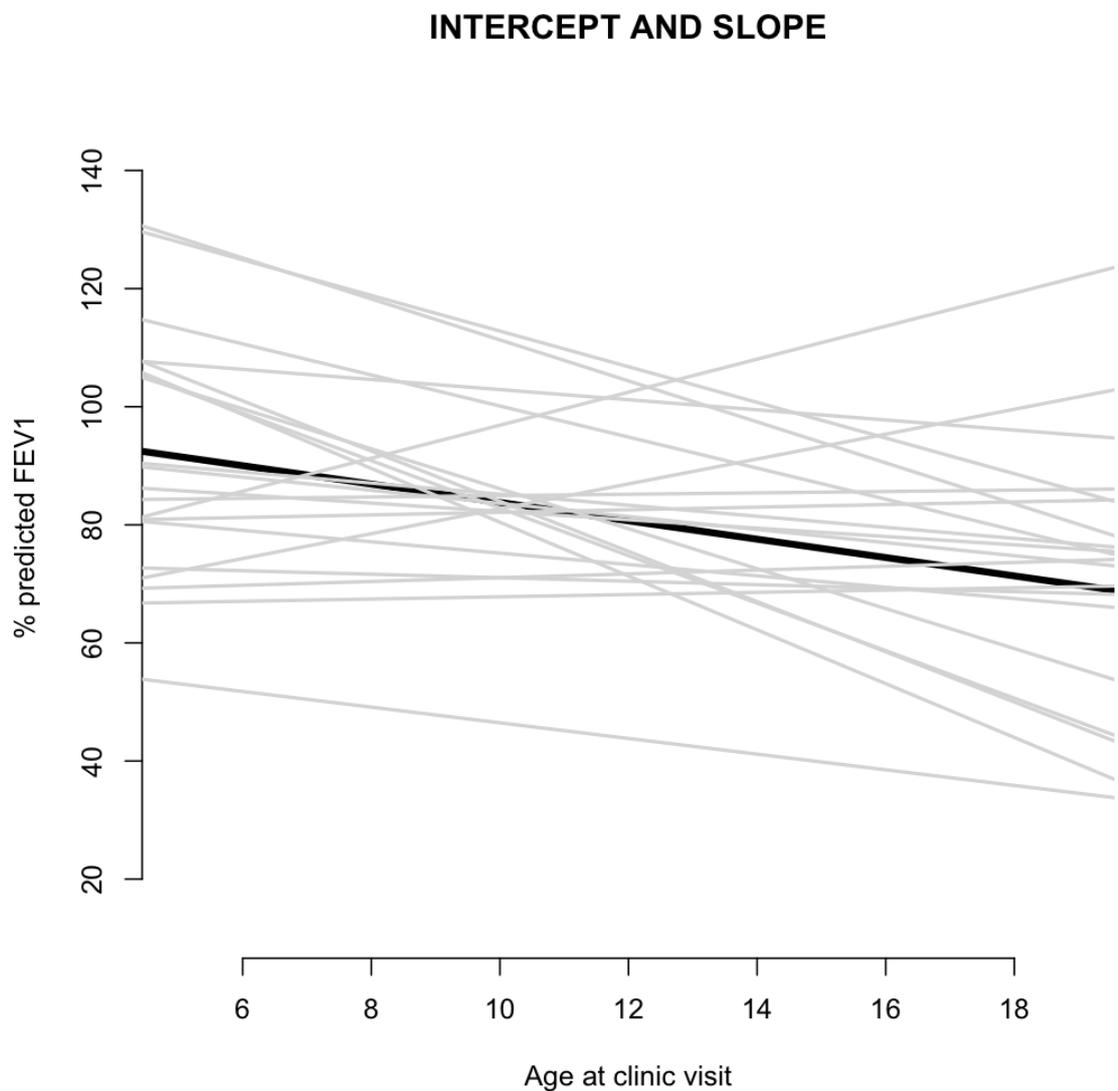
$$\mu_{ij} = \alpha + \sum_{k=1}^p x_{ijk} \beta_k \quad (2)$$

In (2), the  $x_{ijk}$  can be any measured values, whether time-constant or time-varying; for example, sex or age. Despite the model's title, non-linear effects can also be captured. Polynomial time-trends can be defined by including powers of age amongst the  $x_{ijk}$ . Spline functions can be obtained by including both powers of age and indicator variables at selected time-points, called knots. For example, a model in which  $x_{ij1} = \text{age}$  and  $x_{ij2} = 0$  for age less than 10,  $x_{ij2} = \text{age} - 10$  for age greater than 10, defines a linear spline with a single knot, also called a split-line or broken-stick model, with a change in slope at age 10.

In (1) the  $(U_i, V_i)$  pairs are subject-specific intercepts and slopes, modelled as zero-mean bivariate Normally distributed random variables independently realised for different subjects, with means zero, variances  $\sigma_u^2$  and  $\sigma_v^2$  and correlation  $\rho$ , and the  $Z_{ij}$  are residuals modelled as mutually independent, Normally distributed random variables with mean zero and variance  $\tau^2$ . This special case of the linear mixed model implies that the variance of the  $Y_{ij}$  increases with age,  $t$ , as the quadratic function  $\tau^2 + \sigma_u^2 + \sigma_v^2 t^2$ .

A random intercept and slope model for %FEV<sub>1</sub> for the UK paediatric population is illustrated in Figure 20 below. Here the population average trajectory  $\mu_{ij}$  is the thick black line, and is modelled as a linear combination of regression coefficients. The grey lines represent individual trajectories, with each individual having their own deviation from the population average intercept and slope ( $U_i, V_i$ ). These intercepts and slopes are drawn from a Normal distribution, the characteristics of which are estimated from the model.

**Figure 20: Random intercept and slope model for %FEV<sub>1</sub> decline**



The standard random intercept and slope model approach assumes that any deviation of an individual's trajectory from the population mean is linear in time over the whole of the follow-up period apart from independent random errors. This assumption is reasonable over short time-periods, but over longer time-periods the assumption of quadratic variance inherent in the RIS model means that individual data traces can diverge unrealistically. The increasing variance over time can be visualised in Figure 20 above, as the grey lines diverge at older ages. Thus this approach was suitable for analysis of the UK data, characterised by many individuals, followed up over relatively short periods of time. However, for the Danish data, characterised by much longer follow-up periods, an extension of the RIS approach was required.

### **Extending the random-intercept model**

To analyse the Danish data, I used a linear mixed-effects model with longitudinally structured correlation (Diggle et al., 2002, Fitzmaurice, 2004), but model random variation in %FEV<sub>1</sub> over time within an individual subject so that the strength of the correlation of the random variation between two values depends on the corresponding time-separation. This allows a flexible specification of the mean response and incorporates three qualitatively different components of stochastic variation about the mean response (Diggle et al., 2002, Fitzmaurice, 2004). The model thus decomposes the overall random variation in the data into three key components: between subjects; between times within subjects; and measurement error.

The approach follows a number of steps. Firstly, I fitted a provisional model for the mean response by ordinary least squares (OLS) and used the empirical variogram of the residuals to provide initial estimates for the three components of variation, and for the shape of the correlation function of the between-times-within-subjects component. I then re-estimated all of the model parameters by maximum likelihood estimation, and used generalized likelihood ratio statistics to compare nested models, and Wald statistics to test hypotheses about model parameters. I then assessed associations between single or multiple covariates and the population mean %FEV<sub>1</sub> over time, and explore. The technical details follow.

I used a linear mixed-effects model with longitudinally structured correlation. Let  $Y_{ij}$  denote the  $j$ th repeated measurement (here, %FEV<sub>1</sub>) on the  $i$ th patient, and write

$$Y_{ij} = \mu_{ij} + R_{ij} \quad (3)$$

where  $\mu_{ij}$  is the mean, population-averaged, response and  $R_{ij}$  is the stochastic variation about the mean response. The mean response is specified as a linear combination of explanatory variables as per (1) above.

To complete the model-specification we decompose the stochastic term  $R_{ij}$  in (3) into three components, hence

$$R_{ij} = U_i + W_i(t_{ij}) + Z_{ij} \quad (4)$$

where  $t_{ij}$  is the  $j$ th measurement time for the  $i$ th patient and the three components of  $R_{ij}$  are specified as follows. Firstly,  $U_i$  describes how the average lung function of the  $i$ th patient varies about the population-averaged response for all patients with the same values of the explanatory variables  $x_{ijk}$ , for example all males aged 20 years. The model assumes that the  $U_i$  are independent copies of a Normally distributed random variable with mean zero and variance  $v^2$ . Secondly, the stochastic process  $W_i(t)$  describes how the actual lung function of the  $i$ th patient varies over time. The model assumes that the  $W_i(t)$  are independent copies of a stationary Gaussian process with mean zero, variance  $\sigma^2$  and correlation function  $\rho(u) = \text{Corr}\{W_i(t), W_i(t-u)\}$  (Fitzmaurice, 2004). Typically,  $\rho(u)$  decays towards zero as  $u$  increases. In study 3, I used an exponential correlation function,  $\rho(u) = \exp(-|u|/\varphi)$ , in which the parameter  $\varphi$  describes the rate at which the correlation decays towards zero with increasing time-separation,  $u$ . The exponential correlation function is a special case of the Matérn family, which includes a second parameter that allows the correlation function  $\rho(u)$  to assume different shapes if the exponential model does not give a good fit (Matern, 1960). Thirdly,  $Z_{ij}$  describes how the imperfectly measured lung function of the  $i$ th patient at their  $j$ th measurement time,  $t_{ij}$ , differs from their underlying actual lung function, i.e. measurement error. In principle, the properties of the measurement error could be estimated directly by repeated measurement of %FEV<sub>1</sub> within a single follow-up session. In practice, the  $Z_{ij}$  represent the sum of

two sources of variation: pure measurement error and within-patient variation in lung-function on shorter time-scales than the shortest time-interval between successive measurement times,  $t_{ij}$  and  $t_{i,j+1}$ . The model assumes that the  $Z_{ij}$  are independent copies of a Normally distributed random variable with mean zero and variance  $\tau^2$ . The model also assumes that the decomposition of the covariance structure of the  $R_{ij}$ , as shown in (3), is common to all individuals.

### Exploratory analysis for the extended model

Exploratory analysis again consists of identifying a suitable form for the set of mean responses  $\mu_{ij}$  and secondly obtaining initial estimates of the parameters in the model for the stochastic terms  $R_{ij}$ .

For the first of these tasks, I used a combination of ordinary least squares fitting of a regression model, and kernel smoothing. Ordinary least squares gives unbiased estimates of baseline explanatory variable effects whatever the structure of the  $R_{ij}$ , whilst kernel smoothing allows the investigation of possibly non-linear time-trends after adjustment for baseline effects. A kernel smoother is an estimate of the form

$$s(t) = \sum_i \sum_j r_{ij} w_{ij}$$

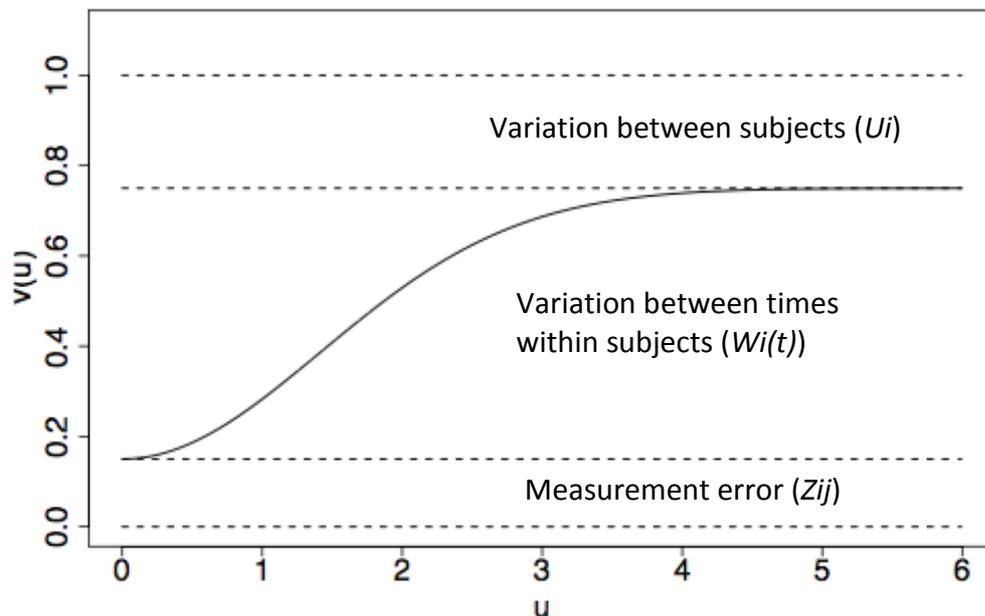
in which the  $r_{ij}$  are the residuals from the regression on baseline explanatory variables whilst the smoothing weights  $w_{ij}$  are scaled to add to 1 and are proportional to  $f(t - t_{ij})$ , where the kernel function,  $f(u)$ , is a probability density function symmetric about  $u = 0$ ; a common choice is a Normal probability density function with mean zero and standard deviation  $h$ . In exploratory analysis, the value of  $h$  can be chosen subjectively so as to obtain a smoothly varying estimate  $s(t)$ .

For the second task, I first re-defined the residuals  $r_{ij}$  to adjust for the estimated smooth time-trend  $s(t)$ . To estimate the covariance structure of these residuals we use the variogram, whose definition is as follows. Let  $v_{ijk} = (r_{ij} - r_{ik})^2/2$  and  $u_{ijk} = |t_{ij} - t_{ik}|$ . Pick a grouping interval  $h$ , let  $n_r$  be the number of  $u_{ijk}$  that lie between  $(r - 1)h$  and  $rh$  and  $\bar{v}_r$  the sample mean of the corresponding  $v_{ijk}$ . A plot of  $\bar{v}_r$  against  $(r - 0.5)h$  is called the *sample variogram* (Diggle et al., 2002, Taylor-Robinson et al.,

2013a). It estimates the function  $V(u) = \tau^2 + \sigma^2\{1 - \rho(u)\}$ , called the theoretical variogram. The sample variance of the residuals estimates the quantity  $\tau^2 + v^2 + \sigma^2$ . Hence, as illustrated in Figure 21 below, by sketching a smooth curve to fit the sample variogram we can obtain initial estimates of the variance components  $\tau^2$ ,  $v^2$  and  $\sigma^2$ , and of the correlation function  $\rho(u)$ .

**Figure 21: A typical example of a theoretical variogram**

Dashed horizontal lines represent a partitioning of the variance into three components, reading from bottom to top,  $\tau^2 = 0.15$ ,  $\sigma^2 = 0.6$  and  $v^2 = 0.25$ . Solid line is the curve  $\tau^2 + \sigma^2\{1 - \rho(u)\}$ .



### Random-effects models for binary data

To analyse binary outcomes (chronic *P. aeruginosa* status, use of therapies in past year), I used a generalized linear mixed model. This specifies a logistic regression model for the effects of covariates on the probability of, for example, *Pseudomonas* acquisition, but adjusts the standard errors of the regression parameters to take account of the correlation structure of the repeated measurements in the same way as described above for the linear mixed model.

Specifically, denoting by  $Y_{ij}$  the  $j$ th repeated binary outcome on the  $i$ th individual,  $t_{ij}$  the age at the time of measurement and  $p_{ij}$  the probability that  $Y_{ij}=1$ , I assumed that

$\log(p_{ij} / (1 - p_{ij})) = \mu_{ij} + U_i + V_i t_{ij}$ , where the  $\mu_{ij}$  are described by a multiple linear regression model and the  $(U_i, V_i)$  pairs are subject-specific intercepts and slopes, modelled as zero-mean bivariate Normally distributed random variables independently realised for different subjects, with means zero, variances  $\sigma_u^2$  and  $\sigma_v^2$  and correlation  $\rho$ .

## **Inference**

I estimated all of the model parameters by maximum likelihood (ML) estimation. This is a popular approach to statistical estimation, due to its favourable properties as sample size increases. ML estimates are asymptotically unbiased, which means they converge on the true value as the sample size increases; they are asymptotically normally distributed, with a variance that can be calculated from the model; and they are efficient, in that ML estimates generate standard errors that are smaller than those obtained through other methods of estimation. Furthermore, functions of ML estimates are themselves ML estimates, which means that the trajectories plotted in this thesis are themselves estimates of the true trajectories, since they are plotted on the basis of ML estimates from a GLM – an intercept value and a slope for example (Singer and Willett, 2003). Asymptotic assumptions rely on large samples, and while the definition of a large sample can be debated, the two datasets analysed in this thesis are comfortably large enough to defend these assumptions.

Conceptually ML estimates are the values of the unknown population parameters of interests that maximise the probability of observing the data. This requires maximising the likelihood function, which is the joint probability density specified by the model in question. Algorithms are used to search for values that maximise this function or the log of the likelihood function, usually on the basis of starting estimates, such as OLS estimates, or those provided by exploratory analysis, for example the empirical variogram parameters outlined above (Singer and Willett, 2003). This is a computationally intensive process, especially for large datasets like those used in this thesis, but can now be routinely undertaken using modern desktop computers.

Likelihood based methods, specifically generalized likelihood ratio tests, can then be used to compare the fit of nested models (i.e. one is a special case of the other). If  $L_1$

and  $L_0$  denote the maximised values of the log-likelihood for nested models with  $p$  and  $p - q$  parameters, the generalized likelihood ratio test for the goodness-of-fit of the simpler model compares  $D = 2(L_1 - L_0)$  with critical values of the chi-squared distribution on  $q$  degrees of freedom. Furthermore, to test hypotheses about model parameters, we use Wald tests. These exploit the property that the maximum likelihood estimates are approximately unbiased and Normally distributed, with standard errors that can be computed from the fitted model; for the full algebraic details, see Diggle, Heagerty, Liang and Zeger (Diggle et al., 2002). Akaike information criterion (AIC) and Bayesian information criterion (BIC) values are also provided in the model tables, and can be applied to non-nested models. For both these measures a lower value suggests a ‘better’ fit (Singer and Willett, 2003).

### **Visualising the results**

I plot the results of the fitted models where appropriate, to demonstrate visually the modelled effect of covariates of interest. For linear mixed effects models this involves plotting the modelled population average (fixed effects) at different values of the variable of interest (e.g. deprivation effect), whilst holding other variables in the model constant. For GLMMs the equivalent calculation needs to take into account the non-linearity of the model. This is achieved by presenting population-averaged percentages or proportions by averaging individual-level fitted values over the population (Diggle et al., 2002).

## Missing values

Some discussion is required of the assumptions regarding missing data implicit in the choice of GLMs for the analyses in this thesis. Missing data, and thus an unbalanced data-frame, is inevitable in large longitudinal observational studies. Missing data on covariates or outcomes in the analyses of interest may occur because patients miss a wave of data collection, owing to a missed appointment, for example, and then they may subsequently appear in the dataset at a later date. Individuals may also drop out of the dataset because they have died, which would be captured in the registry, since deaths are well recorded in this intensively followed-up population, or for unknown reasons.

This missing data is problematic for two main reasons. Firstly, most standard statistical methods assume complete case data, and ignoring missing data can cause serious bias in estimates, if for example the characteristics of people who remain in the dataset are systematically different from those who dropout (Armitage et al., 2008). Secondly dropout may also result in a loss of information and statistical power, especially if a complete case approach is adopted (Diggle et al., 2002, Singer and Willett, 2003). The best way to avoid these issues is to plan to reduce ‘missingness’ at the outset through well-designed experiments and surveys. Since CF is a serious illness, and patients are followed up in a systematic fashion, although the dataset is unbalanced with patients attending at regular intervals, most patients will be reviewed in the UK approximately annually. A key issue, however, is how the models used deal with dropout due to death. The language of missing data is informed by the seminal work of Rubin and Little, who describe three main types of missingness mechanisms (Little and Rubin, 1987):

- Missing completely at random (MCAR): here the missingness is independent of the data, and essentially each person’s observed records are assumed to be a random sample of the data from his or her underlying true trajectory. Furthermore the probability that an outcome is missing is the same for each individual.
- Missing at random (MAR): here the missingness depends only on the observed data, whether this is a covariate included in the analysis, or previous

values of the outcome measure. Thus the probability of missingness is the same for individuals, given the covariates in the model.

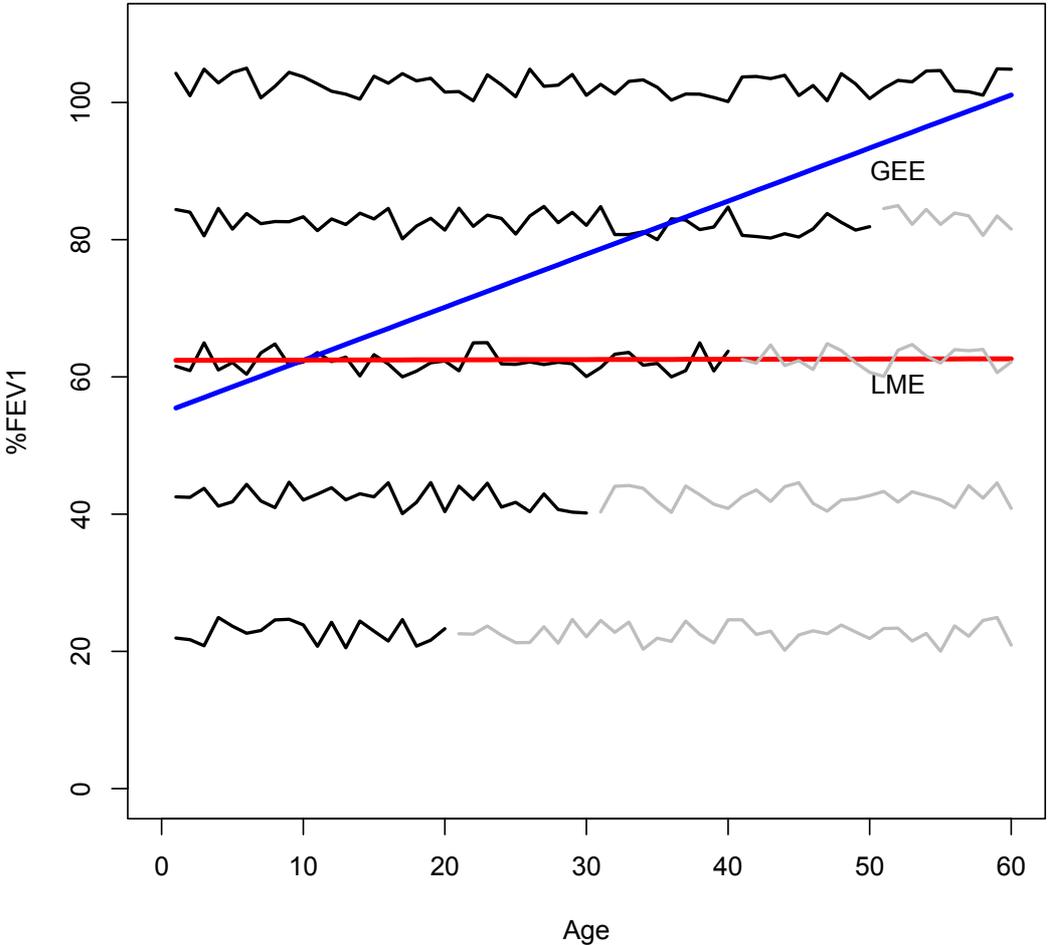
- Not missing at random (NMAR): here the missingness depends on both the observed and missing data.

The assumption of MCAR is difficult to sustain for longitudinal data, since this would imply that the probability of assessment at a particular time is independent of time; the values of the substantive covariates in the model; and the values of the outcome (Singer and Willett, 2003). If the data are NMAR approaches such as selection models, and pattern mixture models are required. The former explicitly incorporates a model for the missingness process, and the latter approach stratifies the analysis on the basis of the observed missingness patterns (Diggle et al., 2002).

The MAR assumption, however, is much less restrictive, and is implicit in the choice of mixed-effects models for the analyses in this thesis. When the data are MAR, the probability of missingness can depend on any of the observed data, either covariates or previous outcomes, and the ML estimates from the GLM will be unbiased estimates, even on the basis of the reduced dataset. Thus in the case of GLM with data assumed to be MAR, the missing values can be classified as ignorable, as long as important correlates of missingness are included in the model. By contrast, the assumptions for a marginal analysis with GEE are generally only consistent under the assumption of MCAR (Diggle et al., 2002).

This is illustrated in Figure 22 below, where I have simulated %FEV<sub>1</sub> data on five individuals, as illustrated in the black lines. Here, dropout from the dataset can be assumed to be due to death, where the risk of death is a function of an individual's level of %FEV<sub>1</sub>, with individuals with lower lung function dropping out earlier.

**Figure 22: Simulated data illustrating the unbiased estimates resulting from a mixed model analysis, where the data is assumed to be MAR, compared to the GEE estimate**



The red line is the population average as fitted using a mixed-effects model, and one can see that this is consistent with the estimate that would have been observed in the absence of dropout i.e. if fitted to the black and grey data. The missing data here can be considered MAR, since the risk death (missingness) is related to covariates in the model e.g. observed values of %FEV<sub>1</sub>. One can see, however, that the population level GEE analysis estimate in blue suggests that in this population of patients, lung function is improving over time, whereas this is entirely the result of selective dropout.

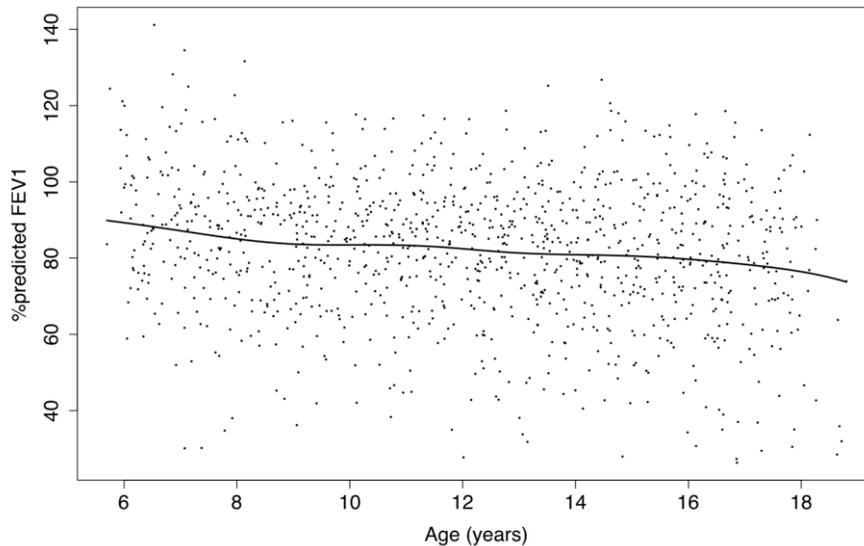
## **Diagnostics**

To test the overall goodness-of-fit of the final models, I analysed the residuals. Plots of residuals against fitted values should show random scatter. Plots of residuals against observed predictor values can also be used to assess patterns in residual variability. Index plots of residuals can be used to identify extreme values in binary models (Singer and Willett, 2003). For the extended model used to analyse the Danish data, the residuals should have approximately the same covariance structure as the fitted model, which I checked by comparing their sample variogram with the theoretical variogram of the model (Diggle et al., 2002).

## **Pilot study and consideration of sample size**

Prior to accessing the UK CF Registry data, I confirmed the feasibility of using the UK CF Register for the studies in this thesis by cleaning and analysing data from two CF treatment centres with good coverage of postcode data (7713 observations on 1423 patients in Liverpool and Belfast). Visual inspection and initial longitudinal analysis of the association between age and % predicted FEV<sub>1</sub> showed the expected statistically significant decline in FEV<sub>1</sub> as age increases, in line with previously published literature (Konstan et al., 2007a). Figure 23 shows the decline in %FEV<sub>1</sub> with age in a paediatric population from one centre. In a random-intercept and slope analysis the regression estimate equated to a fall of 0.58% per year (95%CI 0.29 to 0.87).

**Figure 23: Scatterplot of % predicted FEV<sub>1</sub> versus age with smoothed mean**



The dataset subset allowed estimation of the between-subject (vb) and within-subject (vw) components of variance for %FEV<sub>1</sub> (vb = 306.25 and vw = 87.61). Assuming a common correlation between any two measurements on the same subjects, the estimated standard error of a mean response calculated from n subjects, each of whom provides m measurements, is  $SE(n,m) = \sqrt{(vb+vw/m)/n}$  (Diggle et al., 2002). The study initially envisaged approximately 10,000 subjects providing an average of about six measurements each. For an indicative power/precision calculation, we notionally divide the 10,000 subjects equally into groups of size 2000 according to quintiles of deprivation score. Then, an estimate of the standard error of the overall mean response in each group is  $SE(2000,6) = 0.40$ , and an estimate of the standard error of the difference between the mean responses in any two groups is  $0.40 \times \sqrt{2} = 0.566$ . Hence, a 95% CI for the difference would have width approximately  $0.566 \times 4 = 2.27\%$ .

O'Connor et al report an absolute difference of 5.5% in %predicted FEV<sub>1</sub> between highest and lowest income quintiles (O'Connor et al., 2003). Schechter et al found a difference of 6.7% in %predicted FEV<sub>1</sub> by Medicaid status, which increased to 9.2% with adjustment for various confounders (Schechter et al., 2001). The sample size calculation suggested that we would have sufficient precision to quantify differences such as these. Moreover, the above calculation was likely to prove conservative, for

at least two reasons. First, it treats deprivation as a five-level factor, rather than a continuously varying quantity; if mean outcome varies smoothly with deprivation, a regression analysis would be a more efficient way of estimating how the mean response changes with deprivation. Second, the calculation did not take into account the potential for further gains in efficiency through adjustment for the effects of other variables that are associated with the response.

### **Computation**

All analyses were carried out using the R open-source software environment ([www.r-project.org](http://www.r-project.org)) (version 2.9.2 for mac). Maximum likelihood estimation, generalized likelihood ratio tests and Wald tests used the `lme()` function within the `nlme` package, together with the exponential class of correlation functions. Variogram calculations used a specially written R function developed by Peter Diggle. The R code for this is available in Appendix 6. Other packages used included the `lme4`, `survival`, `Hmisc`, `memisc`, `mcgv` and `ggplot2` packages. Though there were some convergence problems whilst working up some of the more complex models, all of the modelling was successfully undertaken on a MacBook Pro, with some extra memory installed (2.8GHz Intel Core 2 Mac Book Pro with 8GB memory running 64-bit version of R). The final multivariable model for the Danish analysis, for example, would take about 8 hours to converge on this system.

## **Chapter 4: Study 1 – The effect of social deprivation on clinical outcomes and the use of treatments in the UK CF population: A longitudinal study**

### **Abstract**

**Background:** Poorer socio-economic circumstances have been linked with worse outcomes in CF. This study explores, for the first time in a UK-wide cohort, longitudinal weight, height, BMI, %FEV<sub>1</sub>, risk of *Pseudomonas* colonisation, and the use of major CF treatment modalities, and their association with deprivation.

**Methods:** A longitudinal registry study of the UK CF population aged under 40 (8055 people with 49,337 observations between 1996 and 2010). Mixed-effects models were used to assess the association between small area deprivation and clinical and healthcare outcomes, adjusting for clinically important covariates.

**Results:** People from the most deprived areas have significantly lower weight (-0.3 SDs, 95% CI -0.4 to -0.2), height (-0.3 SDs, 95% CI -0.4 to -0.2) and BMI (-0.13 SDs, 95% CI -0.2 to -0.04) in childhood, are more likely to have chronic *Pseudomonas* infection (OR 1.9, 95% CI 1.3 to 2.7), and a significantly lower %FEV<sub>1</sub> (-4.1 percentage points, 95% CI -5 to -3.2). These inequalities are apparent very early in life and do not widen thereafter. On a population level, there is striking evidence of positive discrimination, or ‘pro-poor’ bias towards individuals in the most deprived quintile compared to the least, in provision of IV antibiotics (OR 2.5, 95% CI 1.9 to 3.2) and nutritional therapies (OR 1.78, 95% CI 1.4 to 2.2), after adjusting for disease severity. In contrast, patients from the most disadvantaged areas are less likely to receive DNase or inhaled antibiotic therapy.

**Conclusions:** More disadvantaged children with CF in the UK have significantly worse growth and lung function, but these inequalities do not widen at older ages. There is evidence that, in the NHS, clinicians making decisions about treatments for children, take deprivation as well as disease status into account, and this may mitigate some effects of social disadvantage. Questions remain for the NHS about the provision of therapies such as DNase to people living in disadvantaged areas.

## Introduction

CF is the commonest life-limiting inherited disease among Caucasian populations, with most patients dying prematurely from respiratory failure. Children with CF in the UK and other OECD countries are usually diagnosed in the first year of life (Cystic Fibrosis Trust, 2011), and subsequently require intensive support from family and healthcare services.

CF is of particular interest in the study of mechanisms that generate health inequalities, because, as a classically inherited recessive genetic disease, there is no reason to suspect a social gradient in incidence of the condition. Inequalities can develop, however, in the outcomes experienced by people with the disease. For example, people with CF from socio-economically disadvantaged backgrounds die younger than those in more advantaged social positions in the UK and the US (Britton, 1989, Schechter et al., 2001, O'Connor et al., 2003, Barr HL, 2011). The adjusted risk of death was around four times higher in CF patients in the US with Medicaid cover (taken as an indicator of poverty), compared to those without Medicaid cover (Schechter et al., 2001). In England and Wales, Barr and Fogarty demonstrated a higher risk of age of death at an age above the median in more advantaged social classes, a pattern that has persisted for over four decades (Barr HL, 2011). As with other chronic diseases, this social patterning of survival in CF implies that social and environmental factors are playing a major role in influencing outcomes (Ben-Shlomo and Kuh, 2002, Marmot et al., 2010b). Inequalities in access to specialist healthcare may also be important, since in many health care systems provision and utilisation of services decreases with declining income of patients (van Doorslaer et al., 2006, Stirbu et al., 2011) the so-called 'inverse care law' (Hart, 1971).

This analysis sets out to ascertain if there are differential health outcomes on the basis of SES in the UK CF population (see Diderichsen model in the introduction, Figure 6). To gain a better understanding of when and how inequalities in outcomes develop in CF, I undertook a longitudinal registry study to explore the effect of deprivation on growth, nutrition, lung function, risk of *Pseudomonas* colonisation, and the use of major CF treatment modalities in a UK-wide population cohort, in the context of a universal health care system, free at the point of use.

## Methods

An in depth description of the methods used is found in Chapter 3. An overview is provided here.

### *Design, setting and data source*

I undertook a longitudinal retrospective cohort analysis of individuals in the UK CF registry under the age of 40 at last follow-up, with at least one outcome measurement and a valid postal code between 1996 and 2010. The UK CF Registry is supported and co-ordinated by the UK CF Trust (Adler et al., 2008, Cystic Fibrosis Trust, 2011). The Registry is maintained to a high standard of data quality, and is estimated to include nearly all people thought to have CF in the UK population (Mehta et al., 2004) and is therefore ideally suited to the study of outcomes and treatments across the whole socio-economic spectrum in the UK.

### *Primary outcome and covariates*

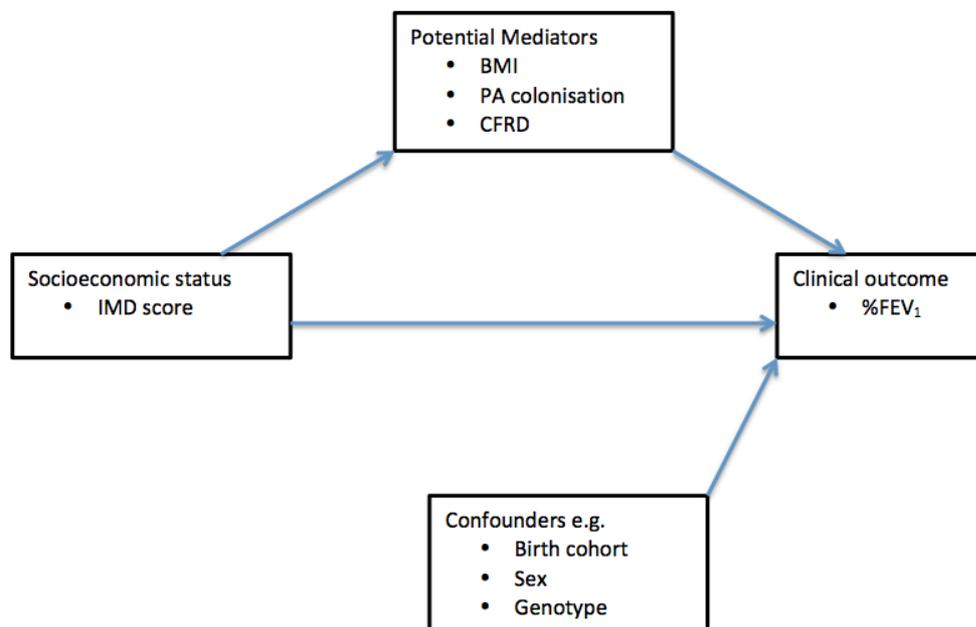
The primary clinical outcomes were weight, height, BMI, %FEV<sub>1</sub> and *Pseudomonas* colonisation prevalence. Anthropometric values were converted into standard deviation scores using the UK reference population (Pan and Cole, 2002). The primary health care outcomes were use of therapies in the previous year (yes or no): IV antibiotics; supplemental nutritional support; DNase; and inhaled antibiotic therapy. Conditional on use of IV therapy, I also used the log total number of days on IV therapy as an outcome.

The primary exposure measure was a small-area-based measure of deprivation of area of residence. Postcodes were used to derive IMD scores for the constituent UK countries (GeoConvert, 2011) and each person was allocated a deprivation score on the basis of the first recorded postcode on entry to the dataset. Other baseline covariates in the analysis were: sex; genotype coded as the number of delta F508 alleles (0, 1 or 2); year of birth; screening status (diagnosis by neonatal screening or otherwise) and ethnicity (Caucasian or otherwise). Time varying covariates were age, presence of CFRD and PI, determined by use of pancreatic enzyme supplementation. In the health care use analyses, I adjusted for disease severity on the basis of current %FEV<sub>1</sub>, *P aeruginosa* status and BMI SD score.

### ***Statistical Methods***

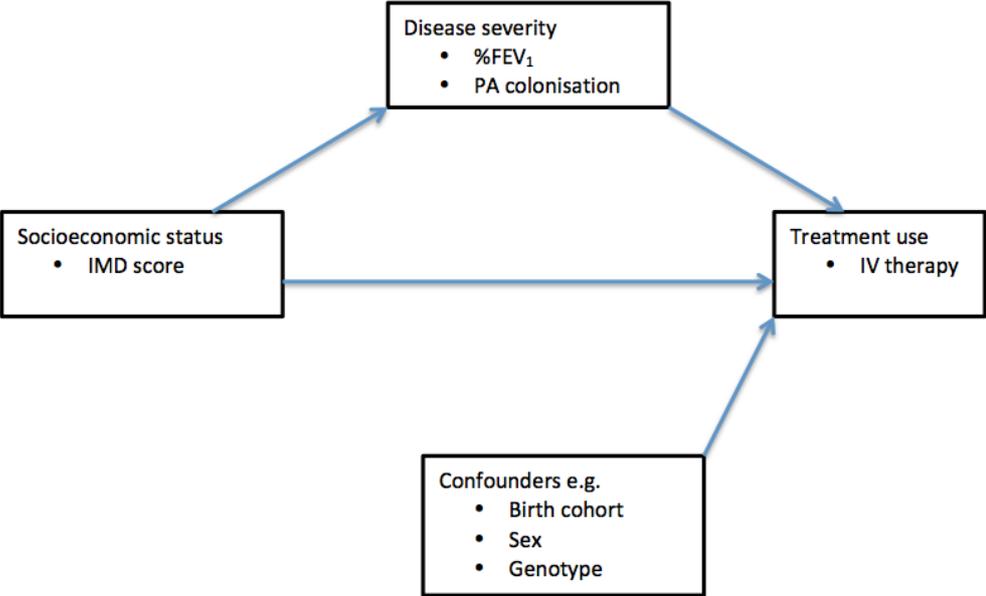
I fitted separate random intercept and slope models in the paediatric (<18) and adult age ranges (18-40). I approximated time-trends using linear functions (e.g. for %FEV<sub>1</sub>), ‘piecewise’ or ‘broken-stick’ functions (weight, BMI), or quadratics (e.g. any IV therapy) as appropriate. For instance, population weight Z score increased to around age three, and then decreased subsequently. The modelling approach involved first fitting a model adjusted for age and the baseline covariates defined above, and then testing for the significance of adding deprivation. Finally, the time-varying covariates were added to the model, to determine if the deprivation coefficient was modified. This is illustrated for the %FEV<sub>1</sub> analysis in Figure 24. The headline results quoted for the effect of SES on outcomes are from the models adjusted for baseline factors that are not plausibly in the causal chain from SES to the outcome of interest. Additional analyses explore the potentially mediating effect of other covariates of interest.

**Figure 24: Logic model for %FEV<sub>1</sub> analysis**



For the analyses of treatment outcomes, I first fitted a baseline model adjusted for baseline factors, and also for disease severity. Then I added deprivation score to the model to ascertain if this has an additional significant effect (Figure 25).

**Figure 25: Logic model for IV therapy analysis**



I estimated all model parameters by maximum likelihood, using linear or generalized linear mixed effects models (Laird and Ware, 1982) and used generalized likelihood ratio statistics to compare nested models, and Wald statistics to test hypotheses about model parameters. Kaplan-Meier estimates and Cox regression were used to assess the effect of deprivation on time to diagnosis.

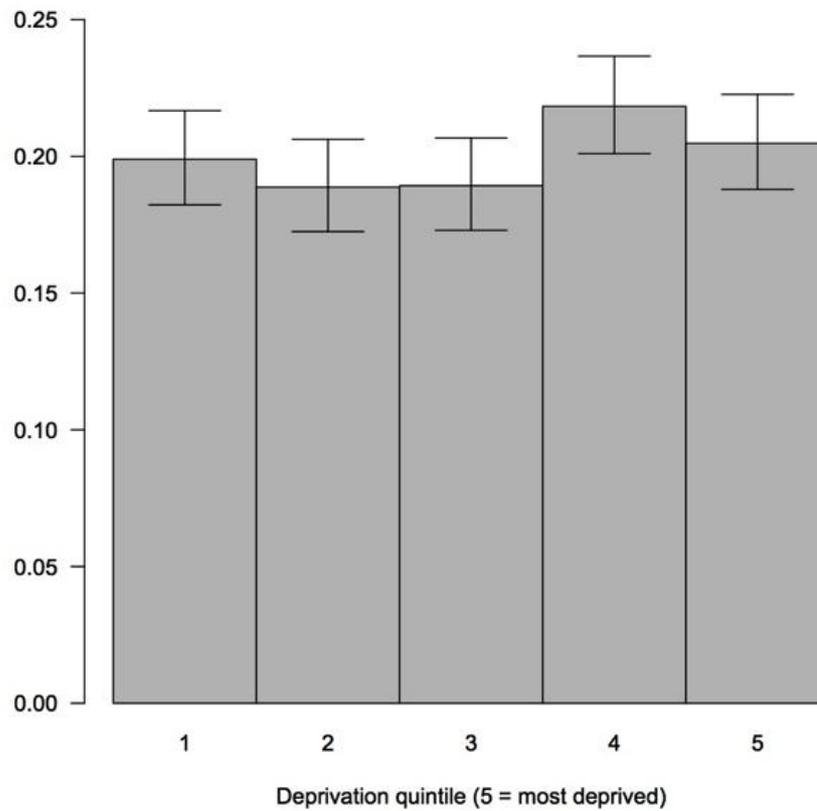
**Results**

**Population characteristics**

The final dataset for weight, the most commonly collected outcome, contains information collected at 49,337 annual reviews on 8055 patients between 1996 and 2010 in the UK. This is described in more detail in the methods chapter. 66% of individuals had five or more follow up measures, with a total of 48,425 person-years of follow up. There was a shallow gradient in the total proportion of individuals by deprivation quintile (19.1% in least deprived versus 20.2 in most deprived) (Table 3), but there was no gradient in incident cases (Figure 26), or number of observations per person (Table 3).

**Figure 26: Distribution of incident cases by deprivation quintile.**

*Error bars represent 95% binomial CIs (n=2066)*



There was no difference in age at diagnosis combining screened and symptomatic diagnoses, sex ratio, birth cohort or neonatal screening by deprivation status. There was a shallow trend towards less heterozygote delta F508 carriers, more people with no delta F508 genes, and a greater proportion of non-white patients, with increasing level of deprivation (Table 3).

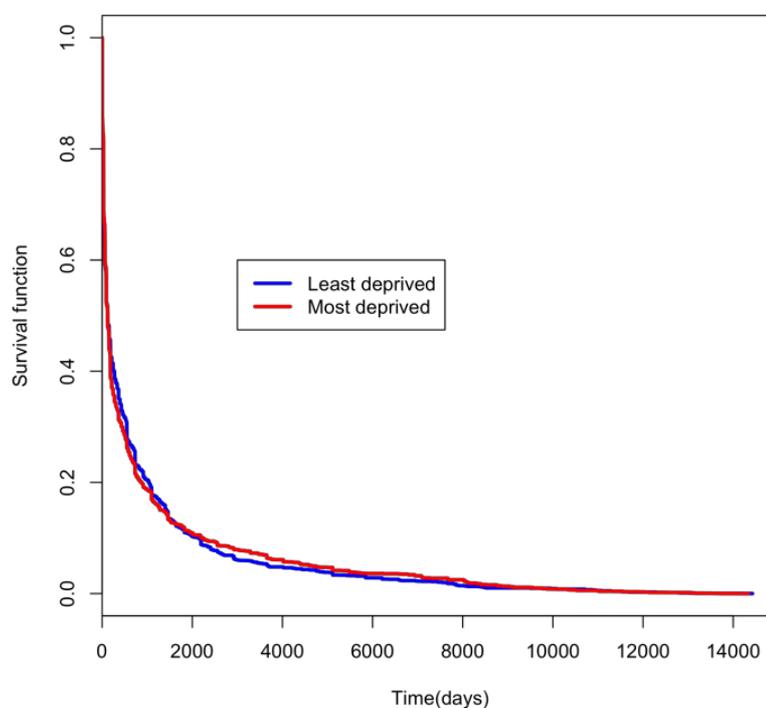
**Table 3: Unadjusted characteristics of study population by deprivation quintile:  
UK CF registry 1996-2010**

	5						
	1	2	3	4	(Most deprived)	All	P
	(Least deprived)						
N (%)	1537 (19.1)	1563 (19.4)	1591 (19.7)	1736 (21.6)	1628 (20.2)	8055	<0.01
Observation (%)	9500 (19.3)	9706 (19.7)	9708 (19.7)	10550 (21.4)	9873 (20.0)	49337	<0.01
Obs per person	6.2	6.2	6.1	6.1	6.1	6.1	0.168
Female (%)	712 (46.3)	726 (46.4)	728 (45.8)	825 (47.5)	773 (47.5)	3764 (46.7)	0.38
Age at diagnosis (days) (IQR)	121 (30,731)	121 (30,669.5)	113 (30,730)	109 (30,727.5)	120 (30,730)	120 (30,730)	
No. delta 508: 2 (%)	824 (53.6)	827 (52.9)	822 (51.7)	907 (52.2)	779 (47.9)	4159 (51.6)	<0.01
No. delta 508: 1 (%)	543 (35.3)	556 (35.6)	560 (35.2)	609 (35.1)	594 (36.5)	2862 (35.5)	0.63
No. delta 508: 0 (%)	170 (11.1)	180 (11.5)	209 (13.1)	220 (12.7)	255 (15.7)	1034 (12.8)	<0.01
Non-white (%)	31 (2)	31 (2)	52 (3.3)	73 (4.2)	120 (7.4)	307 (3.8)	<0.01
Screened (%)	233 (15.2)	272 (17.4)	245 (15.4)	282 (16.2)	277 (17)	1309 (16.3)	0.39
Birth cohort (%)							
>1957-01-01	62 (4)	49 (3.1)	64 (4)	51 (2.9)	35 (2.1)	261 (3.2)	<0.01
>1967-01-01	157 (10.2)	172 (11)	182 (11.4)	171 (9.9)	153 (9.4)	835 (10.4)	0.23
>1977-01-01	329 (21.4)	384 (24.6)	369 (23.2)	426 (24.5)	396 (24.3)	1904 (23.6)	0.09
>1987-01-01	496 (32.3)	478 (30.6)	489 (30.7)	535 (30.8)	530 (32.6)	2528 (31.4)	0.82
>1997-01-01	396 (25.8)	393 (25.1)	396 (24.9)	427 (24.6)	410 (25.2)	2022 (25.1)	0.62
>2007-01-01	97 (6.3)	87 (5.6)	91 (5.7)	126 (7.3)	104 (6.4)	505 (6.3)	0.32

There was no association between deprivation and age at diagnosis in a survival analysis. Figure 27 shows the Kaplan-Meier estimates comparing extremes of deprivation quintile, and Table 4 shows the results of a Cox regression. There was no difference in time to diagnosis comparing the most deprived to the least deprived quintile (hazard ratio (HR) 1.00, 95%CI 0.97 to 1.03), and no consistent trend evident.

**Figure 27: Kaplan-Meier plot of time to diagnosis by deprivation quintile.**

*There is no social gradient in age at diagnosis in the UK CF population.*



**Table 4: Cox regression of time to diagnosis by deprivation quintile**

	$exp(coef)$	$exp(-coef)$	lower .95 CI	upper .95 CI
Quintile 2	1.0534	0.9493	1.0238	1.084
Quintile 3	0.9881	1.012	0.9603	1.017
Quintile 4	1.0032	0.9969	0.9755	1.032
Quintile 5	1.0017	0.9983	0.9737	1.031

(reference quintile 1, least deprived)

## Summary of deprivation effect on outcomes and use of therapies

Before considering each outcome in detail, I provide an overall summary of the key results of this study in Table 5 below. This demonstrates worse clinical outcomes in individuals from the most deprived areas. For IV and nutritional therapies, people in the most deprived quintile are more likely to receive treatment, after adjusting for disease severity. However, for inhaled therapies (DNase and inhaled antibiotics), people in the most deprived quintile are less likely to be treated after adjusting for disease severity.

**Table 5: Summary of adjusted effects of deprivation on clinical outcomes and use of therapies in the paediatric and adult CF population in the UK**

	<i>Age&lt;18</i>	<i>Age 18-40</i>
<b>Clinical outcomes<sup>^</sup></b>		
%FEV <sub>1</sub> (%)	-4.12 (-5.01 to -3.19)	-1.6 (-4.41 to 1.25)
Weight-for-age (SD score)	-0.28 (-0.38 to -0.18)	-0.31 (-0.46 to -0.16)
Height-for-age (SD score)	-0.31 (-0.40 to -0.21)	-0.31 (-0.43 to -0.19)
BMI-for-age (SD score)	-0.13 (-0.22 to -0.04)	-0.12 (-0.25 to 0.01)
OR for <i>P. aeruginosa</i> colonisation	1.89 (1.34 to 2.66)	1.78 (1.26 to 2.51)
<b>Therapies</b>		
OR for any IV therapy*	2.52 (1.92 to 3.17)	1.89 (1.51 to 2.38)
Mean difference IV days per year (%)*	15.9 (8.2 to 24)	10.6 (2.5 to 19.2)
OR for supplemental feeding <sup>§</sup>	1.78 (1.42 to 2.2)	2.38 (1.69 to 3.36)
OR for DNase therapy*	0.40 (0.21 to 0.72)	0.37 (0.26 to 0.52)
OR for inhaled antibiotics <sup>§</sup>	0.66 (0.47 to 0.93)	0.40 (0.31 to 0.5)

All estimates compare the most deprived quintile to the least deprived (reference) quintile. 95% CI in parenthesis

<sup>^</sup>The outcomes are from separate longitudinal models adjusted for time trends, sex, genotype, screening status and ethnicity.

\*Adjusted for time trends, sex, genotype, screening status, %FEV<sub>1</sub> and *P. aeruginosa* colonisation status

<sup>§</sup>Adjusted for time trends, sex, genotype, screening status and BMI SD score

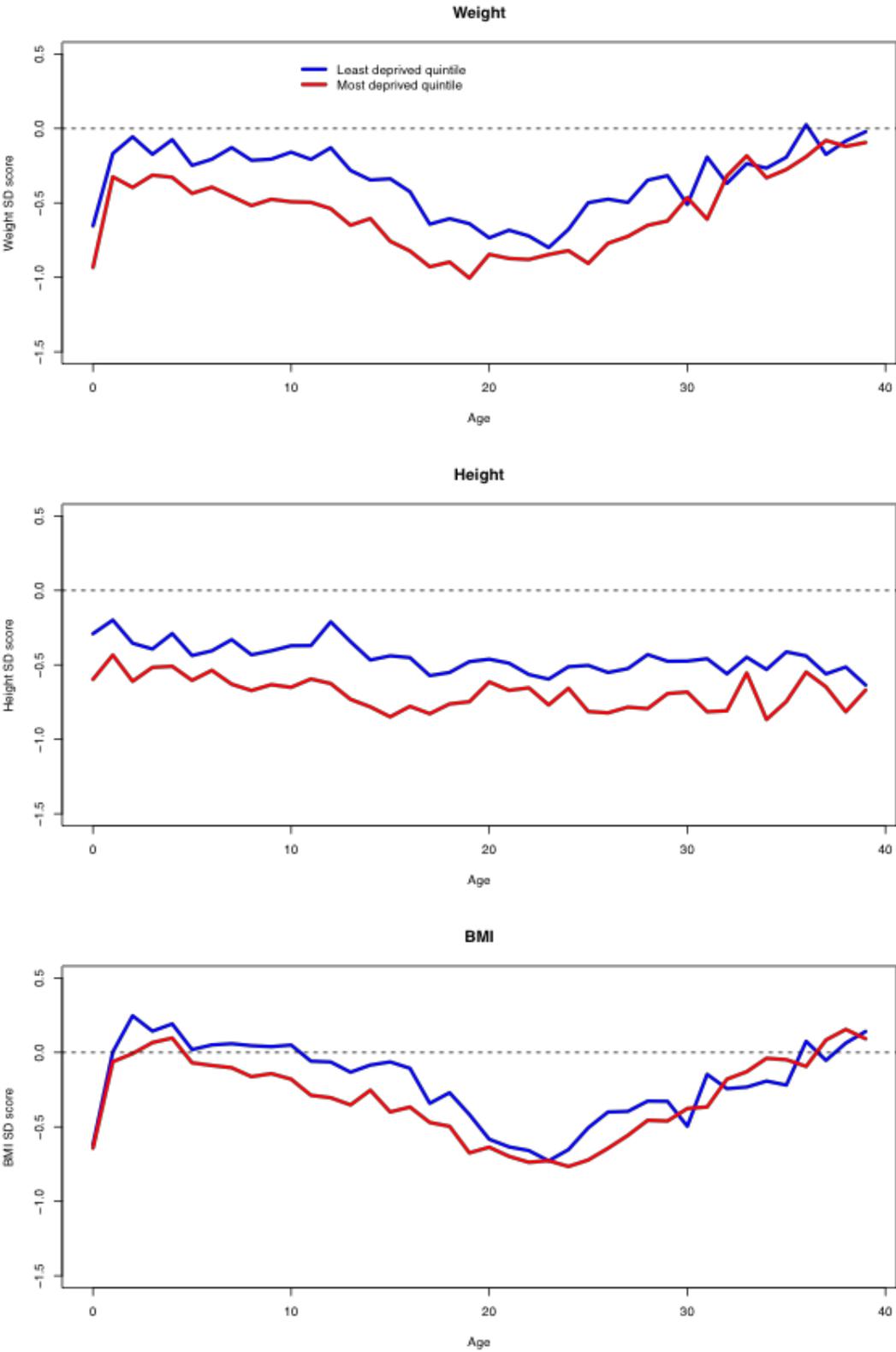
The following sections present the analyses for the clinical outcomes, and treatment outcomes in more detail. The complete analysis is outlined for weight, along with all of the exploratory plots, but for subsequent analyses some of the tables and figures have been placed in the appendix for this chapter (Appendix 4). Robustness tests are in a section at the end of the results.

### **Anthropometric outcomes**

Weight, height and BMI SD scores were significantly lower in the UK CF population compared to the UK reference population in children (-0.37, 95% CI -0.43 to -0.35 [35th centile]; -0.50, 95% CI -0.53 to -0.47 [30th centile]; -0.08, 95% CI -0.11 to -0.06 [46th centile] respectively in longitudinal models ignoring time trends).

First, I provide an overall summary for the three growth outcomes (weight, height and BMI). Figure 28 shows the cross sectional data for each growth outcome by age, stratified by deprivation quintile (Figure 28). People with CF from the most deprived quintile (red line), compared to the least deprived (blue line) in the UK have a lower weight, height and BMI SD scores.

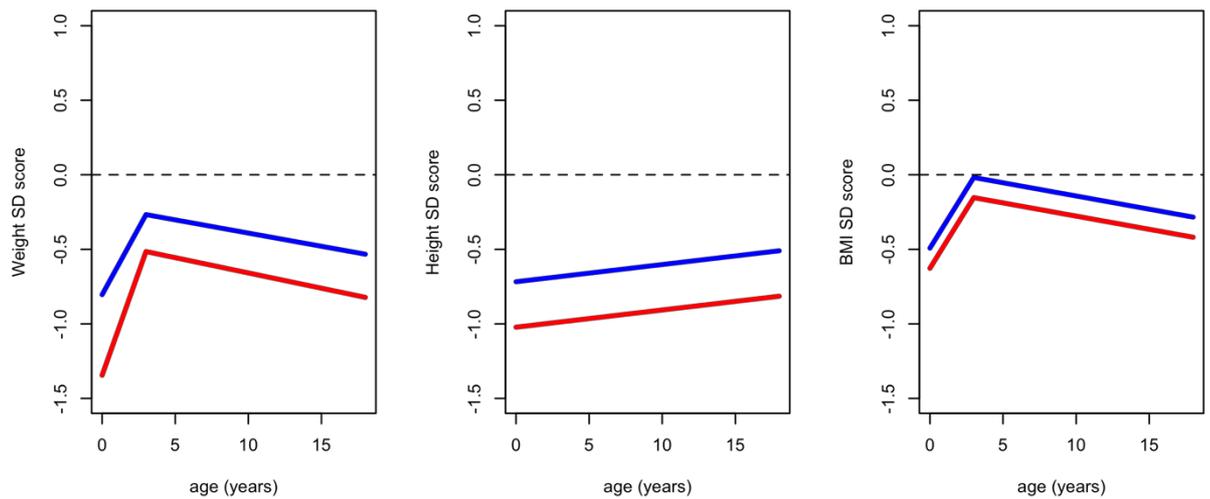
**Figure 28: Anthropometric outcomes: mean cross-sectional weight, height, and BMI by age comparing extremes of deprivation quintile (red most deprived).**



### *Summary of longitudinal analysis*

The plots in Figure 29 illustrate the contrast between deprivation quintiles from the final longitudinal models for the <18 age group. The trajectories are plotted at the reference values for other covariates in the final regression models: female sex, homozygote delta F508 carrier, not diagnosed by screening, white, born in 1991. Weight SD scores increased from the time of diagnosis to around age three, and then decreased. This is modelled as a split straight line with a break point (knot) at age three.

**Figure 29: Modelled growth trajectories for children, comparing least (blue) and most deprived quintiles (red).**

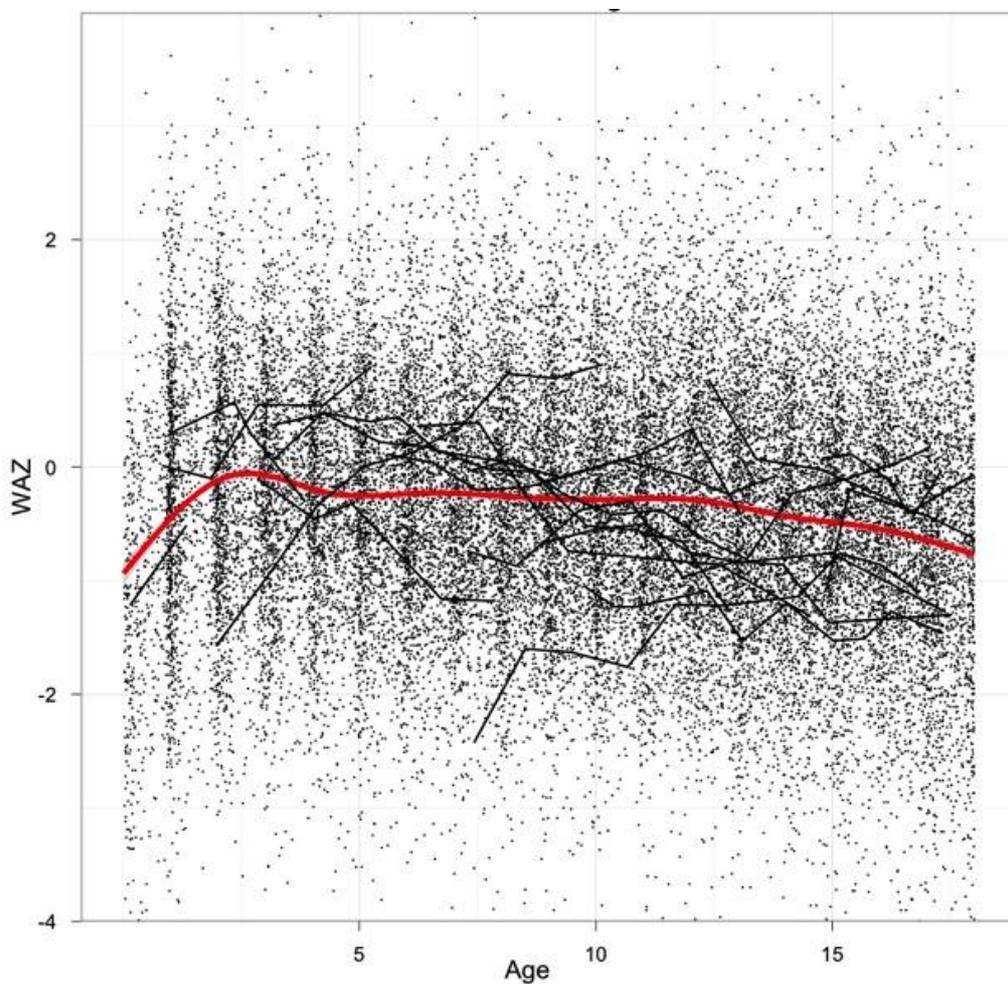


### **Weight**

The mean weight for age Z score (WAZ or weight SD score) for the population over time (null model) was -0.37 (95% CI -0.43 to -0.35), which corresponds to around the 35<sup>th</sup> centile on a growth chart. Exploratory analysis begins with spaghetti plots, with smoothed means added to explore the form of the overall mean trajectory. Figure 30 shows the spaghetti plot for the <18 analysis.

**Figure 30: Spaghetti plot of weight SD score versus age in paediatric age group**

*Mean smoother in red line, and randomly selected weight trajectories in black.*

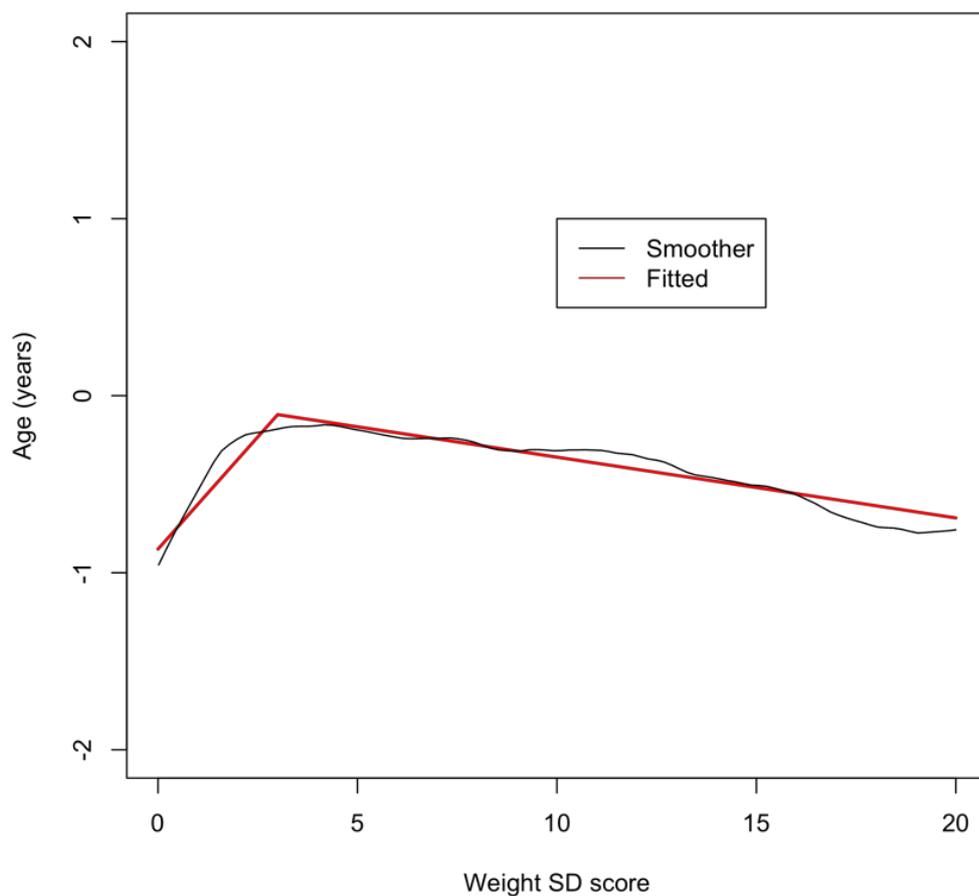


Weight SD score in the population actually increases from the time of diagnosis to around age three, and then decreases subsequently. This is modelled as a split

straight line (piecewise regression) with a knot at age three in the under 18 age group (Figure 31).

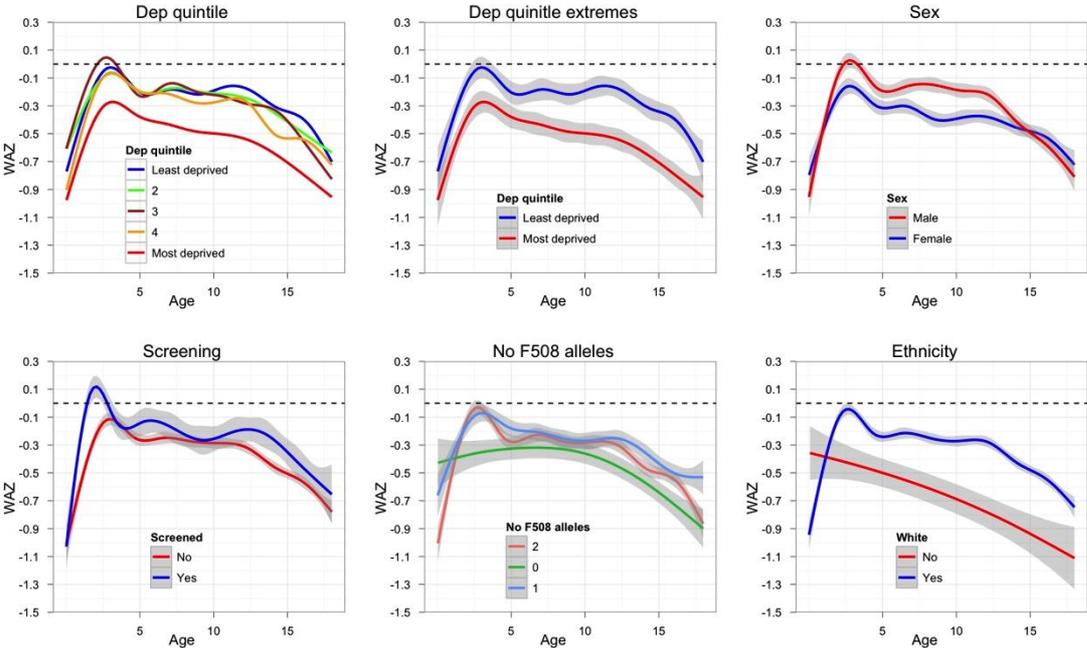
**Figure 31: Piecewise modelling approach to weight z score trajectory**

*The plot shows the broken stick model, fitted by OLS compared to smoothed mean. The smoothed mean weight z score increases to around age three and decreases subsequently. This was modelled as a piecewise regression, with a 'knot' at age three.*



Undertaking exploratory analyses plotting smoothed cross-sectional means for baseline covariates of interest suggested important covariate effects, and also confirmed that the split-line time trends were consistent across covariate subgroups (Figure 32).

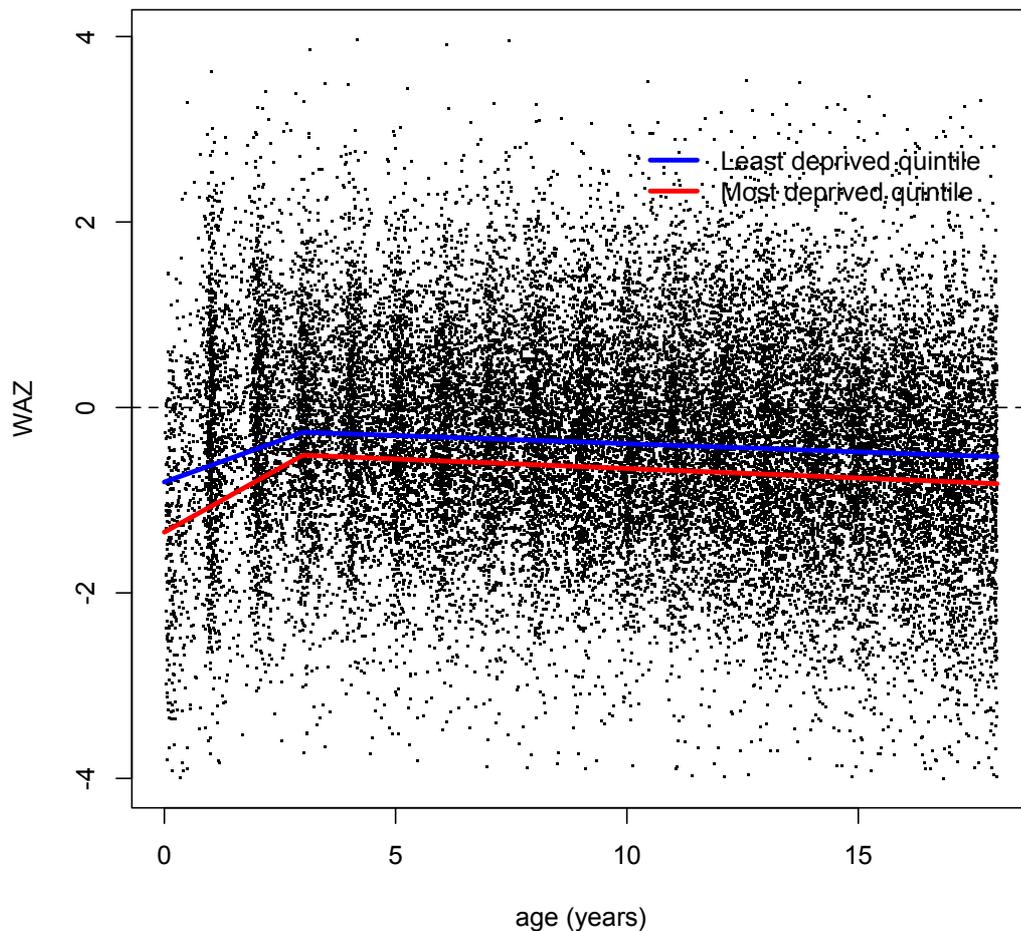
**Figure 32: Exploratory analysis showing smoothed means of weight for age SD score versus age, stratified by covariates, for people <18**



Moving now to the definitive longitudinal analysis for weight SD score in the <18 group, the results of the final model are shown in Table 6. The important covariate effects are visualised in the subsequent plots. After adjustment for baseline factors, the weight of children in the most deprived quintile compared to the least was significantly lower around diagnosis (-0.54, 95% CI -0.73 to -0.34). The deprivation gap narrowed slightly up to age three, and from then on remained constant (-0.28, 95% CI -0.38 to -0.18, comparing most to least deprived) (Figure 33). Addition of the time-varying covariates did not substantially alter the deprivation effects for weight (Appendix 4, Table 20), and the estimates were consistent with a monotonic dose response relationship with deprivation (see robustness tests section).

**Figure 33: Deprivation quintile contrast from final weight model <18**

*Trajectories plotted at reference values for other covariates in the regression model: female sex, homozygote delta F508 carrier, not diagnosed by screening, white, born in 1991.*



**Table 6: Final linear mixed-effects regression models for growth in the <18 age group**

	<i>weight_young</i>	<i>height_young</i>	<i>bmi_young</i>
Constant	-0.76679*** (0.09124)	-0.69634*** (0.07615)	-1.43559* (0.59086)
age	0.17231*** (0.01769)	0.01154*** (0.00296)	0.15798*** (0.02777)
age2	-0.18983*** (0.01838)		-0.17570*** (0.02880)
Number of F508 alleles: 0/2	0.06228 (0.10612)	-0.05825 (0.07068)	-0.03069 (0.15020)
Number of F508 alleles: 1/2	0.01447 (0.06184)	-0.00004 (0.04431)	-0.21288* (0.08857)
Male	0.03527 (0.05724)	0.17969*** (0.04094)	-0.09166 (0.08159)
Non-white	0.33507* (0.14012)	0.23514* (0.10062)	-0.02379 (0.18750)
Screened	0.28633*** (0.06427)	0.27879*** (0.05056)	0.10080 (0.08734)
Deprivation score	-0.00932*** (0.00169)	-0.00526*** (0.00086)	-0.00229** (0.00079)
Age x Number of F508 alleles: 0/2	-0.04279 (0.03088)	0.00184 (0.00529)	0.00141 (0.05455)
Age x Number of F508 alleles: 1/2	-0.02331 (0.01687)	0.00582 (0.00347)	0.05660 (0.03192)
Age x male	0.05831*** (0.01568)	-0.01218*** (0.00331)	0.11384*** (0.02947)
Age x non-white	-0.17607*** (0.04040)	-0.02756*** (0.00804)	-0.12099 (0.07002)
Age x screened	-0.06239*** (0.01666)	-0.01710*** (0.00407)	-0.03067 (0.03215)
Age x Deprivation score	0.00168*** (0.00045)		
age2 x Number of F508 alleles: 0/2	0.04497 (0.03199)		0.00036 (0.05688)
age2 x Number of F508 alleles: 1/2	0.03611* (0.01755)		-0.04850 (0.03332)
age2 x male	-0.07335*** (0.01636)		-0.13381*** (0.03079)
age2 x non-white	0.16692*** (0.04254)		0.13337 (0.07400)
age2 x screened	0.06005*** (0.01769)		0.03464 (0.03407)
Age2 x Deprivation score	-0.00173*** (0.00047)		
Log-likelihood	-25528.81159	-22657.81459	-28158.12899
Deviance	51057.62317	45315.62917	56316.25798
AIC	51167.62317	45409.62917	56428.25798
BIC	51623.19488	45798.52902	56891.62221
N	29235	28983	28980
Groups	5775	5750	5745

\* P < 0.05, \*\* P < 0.01, \*\*\* P < 0.001

Standard errors in parentheses, birthyear coefficients not shown

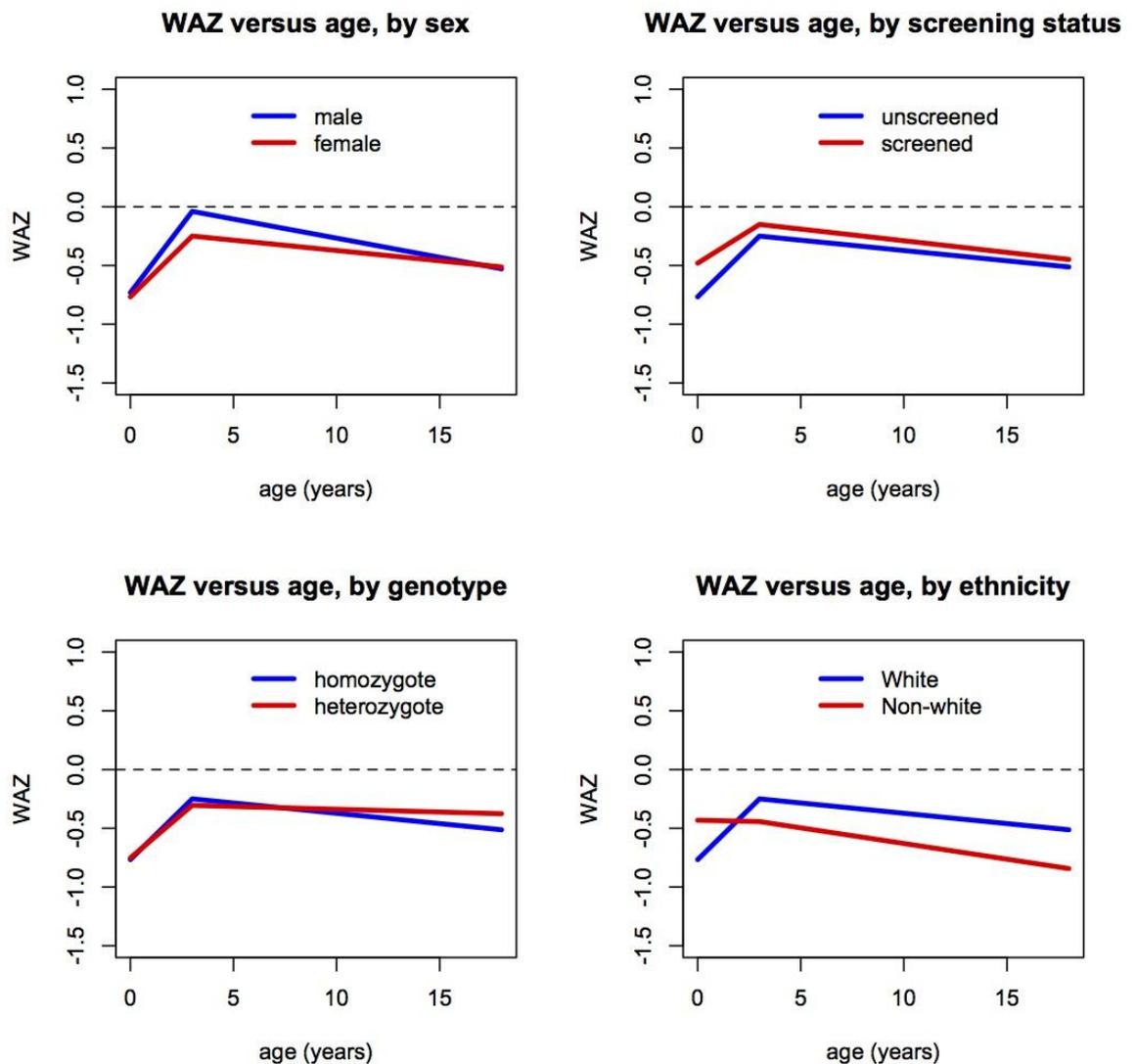
The deprivation effect is multiplied by 58 to generate the contrast between the mid point of the least and most deprived quintiles

age2 is the coefficient for the split line at age three in the weight and BMI analysis

Higher weight SD score was associated with male sex, screened patients, heterozygotes for delta F508 and white patients in the <18 age group (Figure 34).

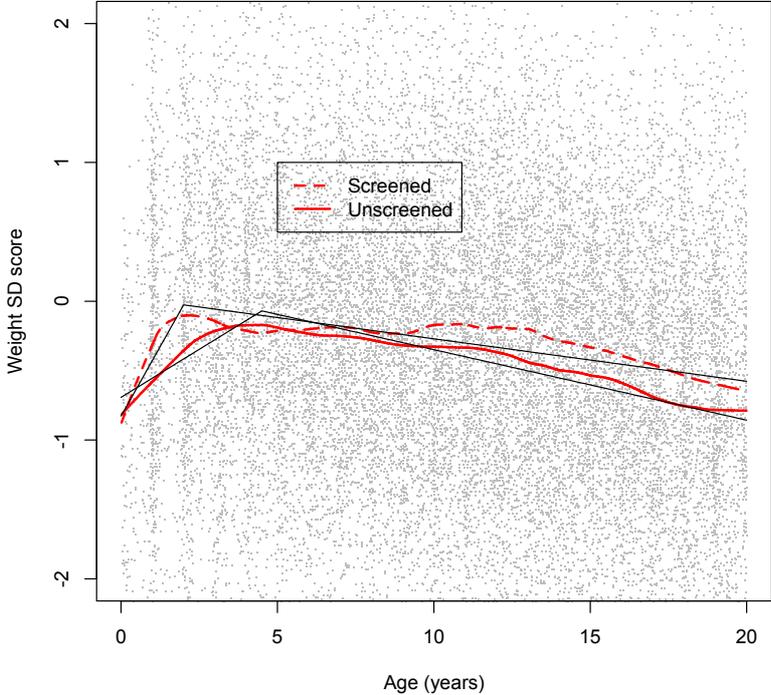
**Figure 34: Weight for age Z score versus age, illustrating the effect of sex, screening status, delta F508 carrier status, and ethnicity**

*Trajectories plotted at reference values for other covariates in the final regression model: female sex, homozygote delta F508 carrier, not diagnosed by screening, white, born in 1991.*



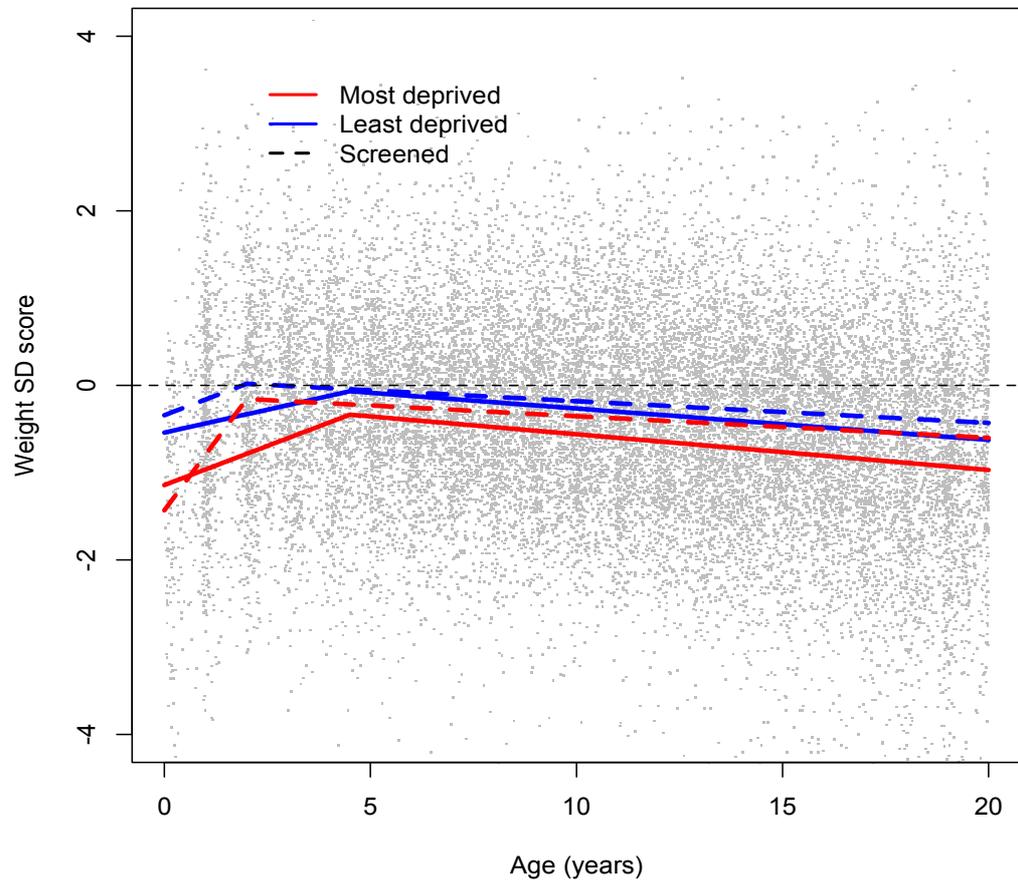
In a supplementary analysis for weight, I stratified the data by screening status, and allowed the break point in the split-line to vary in each group (Figure 35), and re-fitted the longitudinal model separately to the screened and unscreened children. The point estimate for the inequality in weight was narrower in the screened group (Figure 36). I tested for an interaction between deprivation and screening status in the full dataset, and although the point estimate was in the direction that supports a narrowing of inequality with screening, it was not significant.

**Figure 35: Piecewise modelling approach to weight SD score trajectory in analysis stratified by screening status**



**Figure 36: Deprivation effect on weight in screening stratified analysis**

*Screening is associated with a narrowing of inequality in weight*

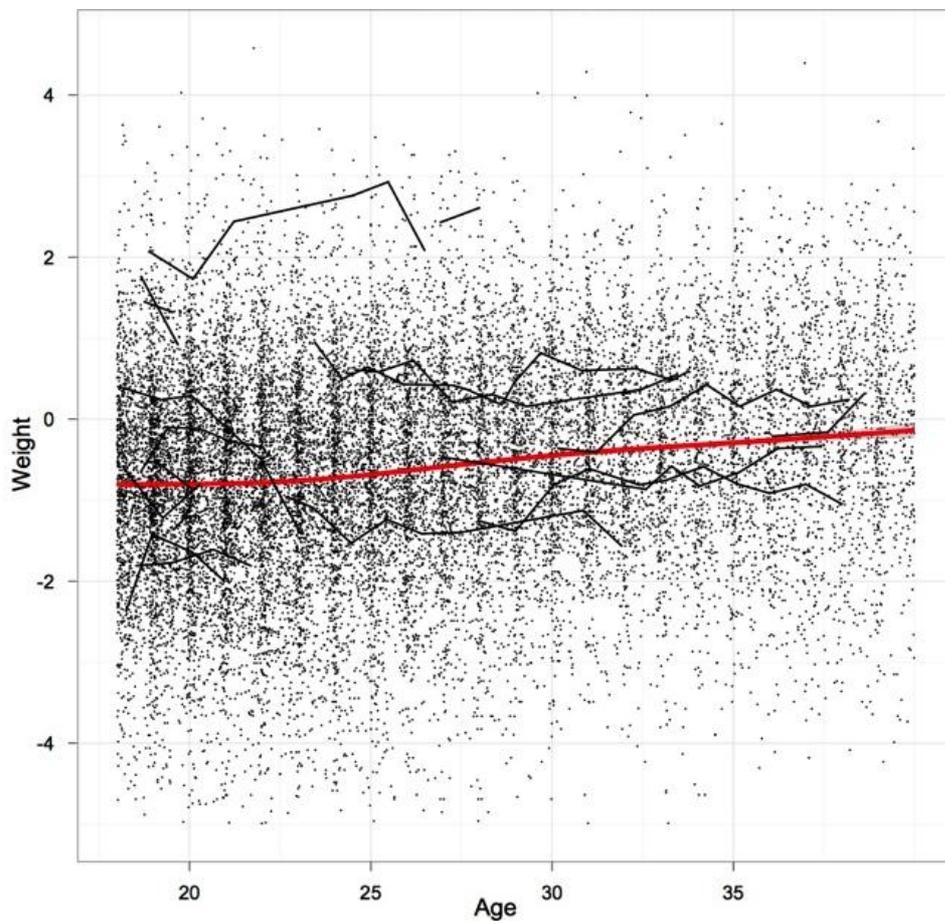


### ***Adult Weight***

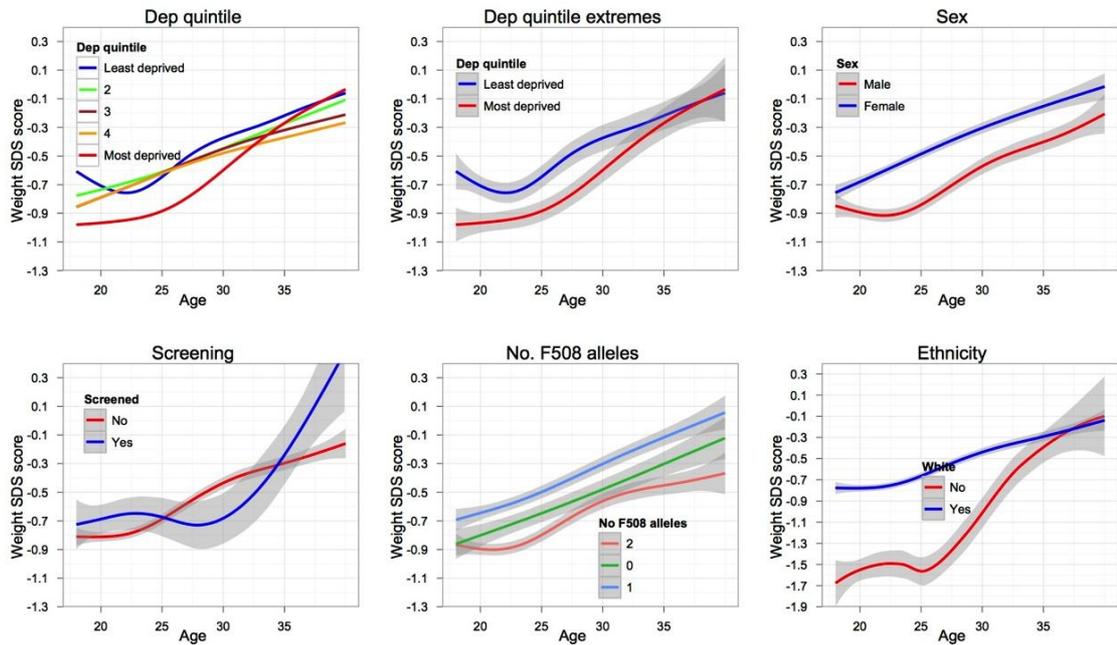
Following the same analysis sequence for weight in the adult age group, the spaghetti plot and exploratory cross-sectional analysis stratified by covariates of interest are shown below (Figure 37, Figure 38). The weight SD trajectory was modelled as a linear function in the adults (Figure 39).

**Figure 37: Spaghetti plot of weight for age SD score versus age in adult age group**

*Mean smoother in red line, and randomly selected weight trajectories in black.*

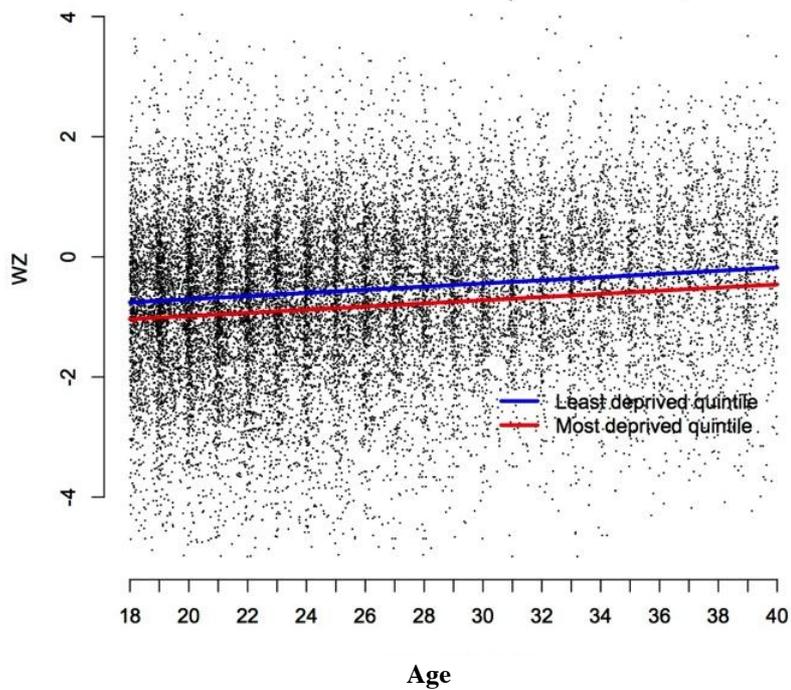


**Figure 38: Exploratory analysis showing smoothed means of weight for age SD score versus age, stratified by covariates, for people >18**



**Figure 39: Deprivation quintile contrast from final weight model >18**

*Trajectories plotted at reference values for other covariates in the regression model.*



The results of the final model for weight SD score are shown for the adult age group in Table 7. In adults the adjusted difference between deprivation quintiles remained constant (-0.31, 95% CI -0.46 to -0.16, comparing most to least deprived) (Figure 39). Addition of the time-varying covariates did not substantially alter the deprivation effects for weight (Appendix 4, Table 21).

**Table 7: Final linear mixed-effects regression models for growth in the >18 age group**

	<i>Weight&gt;18</i>	<i>Height&gt;18</i>	<i>BMI&gt;18</i>
Constant	-0.81353*** (0.10244)	-0.22411** (0.08435)	-0.17173 (0.09519)
Age-18	0.01839*** (0.00391)	-0.00508*** (0.00140)	-0.07677*** (0.01117)
age2			0.10152*** (0.01333)
Number of F508 alleles: 0/2	0.02906 (0.06518)	0.11796** (0.03626)	-0.06111 (0.07592)
Number of F508 alleles: 1/2	0.22908*** (0.04973)	0.10350** (0.03227)	0.18013*** (0.05347)
Male	-0.33225*** (0.04714)	-0.24207*** (0.03474)	0.00446 (0.04934)
Non-white	-0.54262*** (0.14092)	-0.38022*** (0.08558)	-0.27837 (0.15174)
Screened	0.06145 (0.07886)	-0.07743 (0.04337)	0.04777 (0.08260)
Age-18 x Number of F508 alleles: 0/2	0.01236 (0.00633)	-0.00191 (0.00238)	0.02537 (0.02019)
Age-18 x Number of F508 alleles: 1/2	0.00123 (0.00446)	0.00006 (0.00165)	-0.00980 (0.01363)
Age-18 x male	0.01053* (0.00425)	0.00018 (0.00155)	-0.03934** (0.01252)
Age-18 x non-white	0.01385 (0.01431)	-0.00119 (0.00528)	0.03412 (0.03951)
Age-18 x screened	-0.01328	-0.00347	0.02532 (0.02234)
age2 x Number of F508 alleles: 0/2			-0.01423 (0.02384)
age2 x Number of F508 alleles: 1/2			0.01497 (0.01618)
age2 x male			0.06495*** (0.01495)
age2 x non-white			-0.02708 (0.04852)
age2 x screened			-0.05189 (0.02884)
Deprivation score	-0.00537*** (0.00130)	-0.00530*** (0.00105)	-0.00203 (0.00117)
Log-likelihood	-18646.75337	-3710.19476	-19860.93596
Deviance	37293.50675	7420.38951	39721.87192
AIC	37395.50675	7522.38951	39841.87192
BIC	37799.39164	7926.13366	40315.94301
N	20319	20263	19954
Groups	4041	4046	4029

\* P < 0.05, \*\* P < 0.01, \*\*\* P < 0.001

Standard errors in parentheses, birthyear coefficients not shown

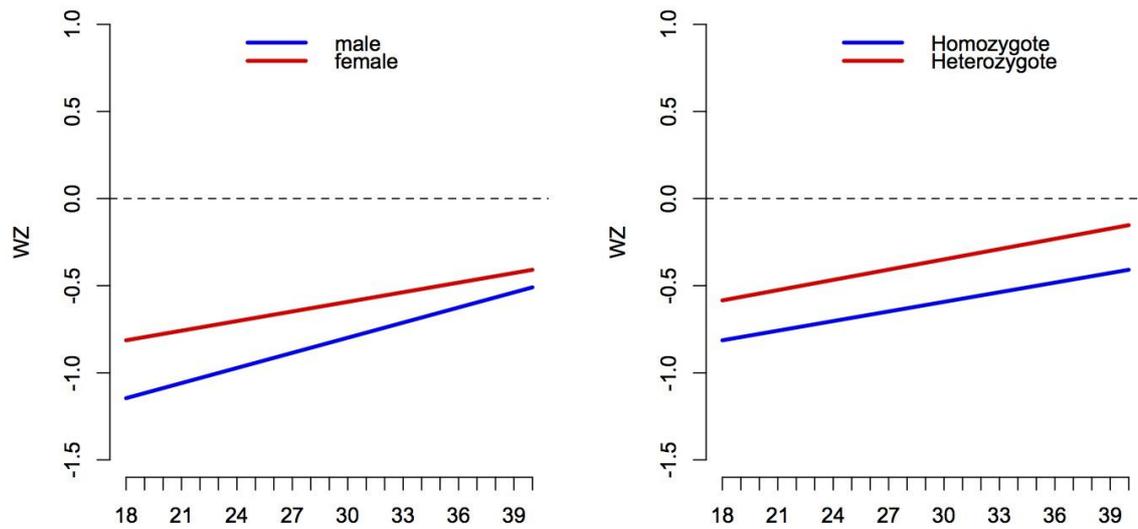
The deprivation effect is multiplied by 58 to generate the contrast between the mid point of the least and most deprived quintiles

age2 is the coefficient for the split line at age three in the BMI analysis

Higher weight SD score was associated with female sex, heterozygote status and white patients (Figure 40) in the >18 age group.

**Figure 40: Weight for age versus age, illustrating covariate contrasts from the final longitudinal models**

*Trajectories plotted at reference values for other covariates in the regression model.*

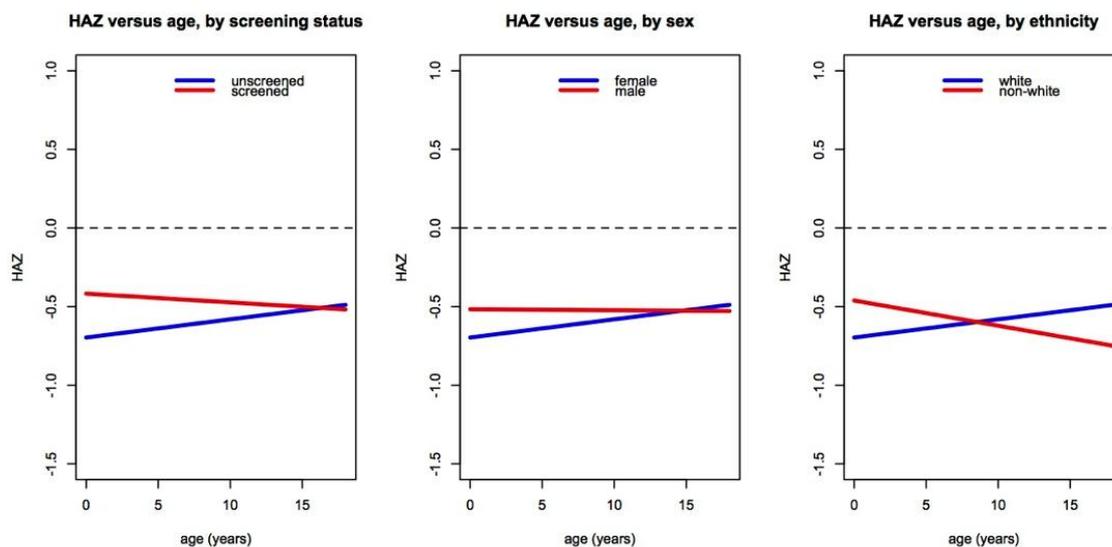


## Height

The mean height for age Z score (HAZ) for the population over time (null model) was  $-0.50$  (95% CI  $-0.53$  to  $-0.47$ ), which corresponds to around the 30th centile on a growth chart. This was modelled as a linear trend in the paediatric and adult age ranges (see exploratory analysis for height in Appendix 4, Figure 83, Figure 84, Figure 85, Figure 86). As for weight, the average height of individuals in the most deprived quintile compared to the least was also about a third of an SD score shorter in the adjusted analysis ( $-0.31$  95%CI  $-0.40$  to  $-0.21$  in the  $<18$  analysis, Figure 29) and this effect remained constant across the paediatric and adult age range (Table 6, Table 7). Greater height SD score was associated with male sex and screened patients over the paediatric age range, and increased in white patients over time (Figure 41). In adults greater height was associated with female sex, non-heterozygotes for delta F508 and white patients (Table 7). Addition of the time-varying covariates did not substantially alter the deprivation effects for height (Appendix 4, Table 22, Table 23), and the estimates were consistent with a monotonic dose response relationship with deprivation (see robustness tests section).

**Figure 41: Height for age versus age, illustrating covariate contrasts from the final longitudinal models**

*Trajectories plotted at reference values for other covariates in the regression model as above.*

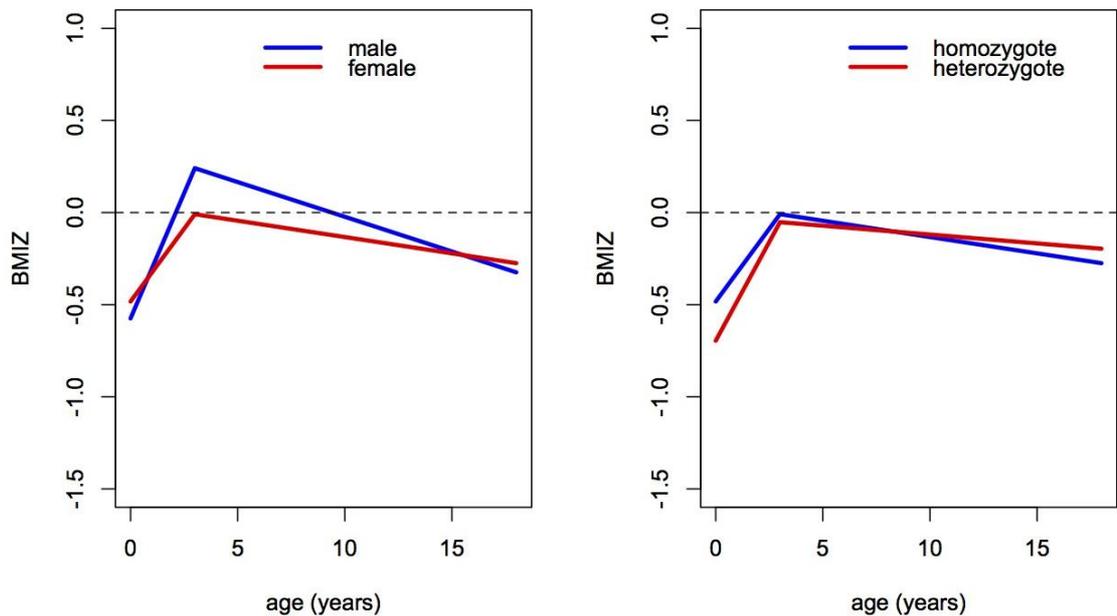


## **BMI**

The mean BMI SD score for the population over time (null model) was -0.08 (95% CI -0.11 to -0.06), which corresponds to around the 46th centile on a growth chart. BMI SD score was modelled in a similar way to weight SD score, with a split line at age three (see Appendix 4 for exploratory analysis, Figure 87, Figure 88, Figure 89, Figure 90). In the paediatric age range there was a deprivation gap of -0.13 (95% CI -0.04 to -0.22), which remained constant over time (Table 6, Figure 29). Higher BMI was associated with male sex in the paediatric age range. BMI SD score had a steeper rate of decline in delta F508 homozygotes after the age of three (Figure 42). The effects of sex and genotype in the adult age group are shown in Figure 43.

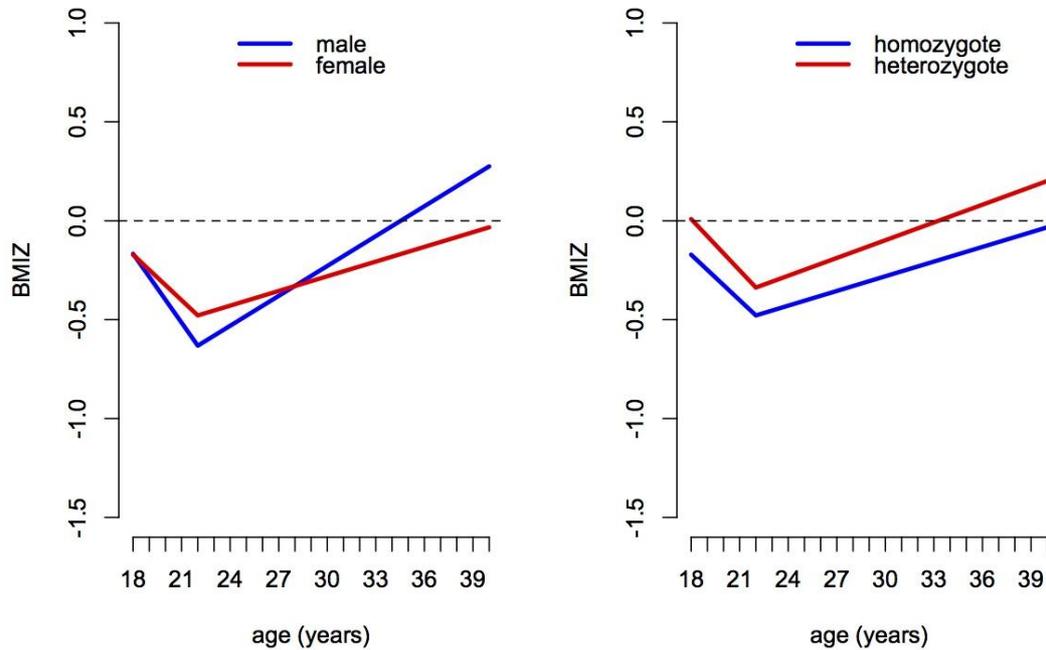
**Figure 42: BMI for age versus age, illustrating covariate contrasts from the final longitudinal models**

*Trajectories plotted at reference values for other covariates in the regression model as above.*



**Figure 43: BMI for age versus age, illustrating covariate contrasts from the final longitudinal models**

*Trajectories plotted at reference values for other covariates in the regression model.*

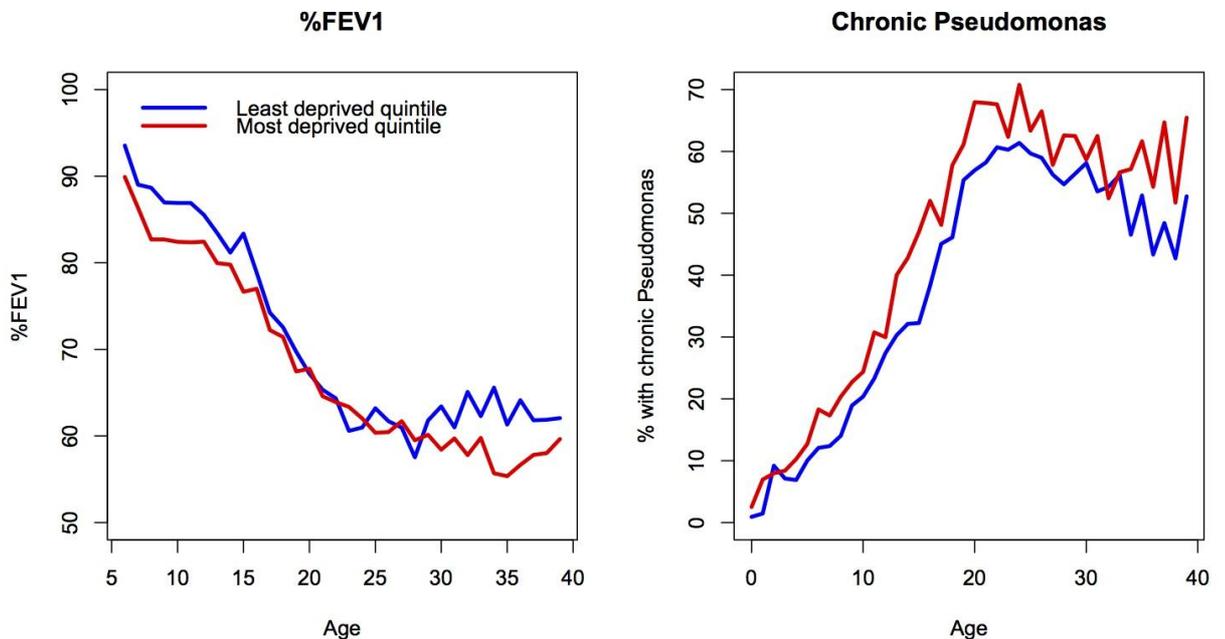


In the adult age range the point estimate for the deprivation effect on BMI SD score was similar to that in the paediatric age group, but this did not reach significance at the conventional 5% level (-0.12, 95% CI -0.25 to 0.01). Addition of time-varying covariates did not substantially alter the deprivation effects for BMI (Appendix 4, Table 24, Table 25).

## Respiratory outcomes

First, I provide an overall summary for the two respiratory outcomes (% FEV<sub>1</sub> and *P. aeruginosa*). Figure 44 shows the cross sectional data by age, stratified by deprivation quintile. People with CF from the most deprived small areas in the UK have a lower %FEV<sub>1</sub> in the <18 age group, and are more likely to be chronically colonized with *P. aeruginosa*. The social gradient in %FEV<sub>1</sub> (around 4% between the most and the least deprived quintiles) in the UK population is already present as soon as children are able to perform spirometry at around the age of five, and does not widen further over the paediatric age range. The prevalence of chronic *P. aeruginosa* infection was greater in the most deprived quintile compared to the least (OR 1.9), after adjustment for baseline covariates, and this association remained after adjustment for level of %FEV<sub>1</sub>.

**Figure 44: Respiratory outcomes: Mean cross-sectional %FEV<sub>1</sub> and *P. aeruginosa* colonization prevalence by age comparing extremes of deprivation quintile**



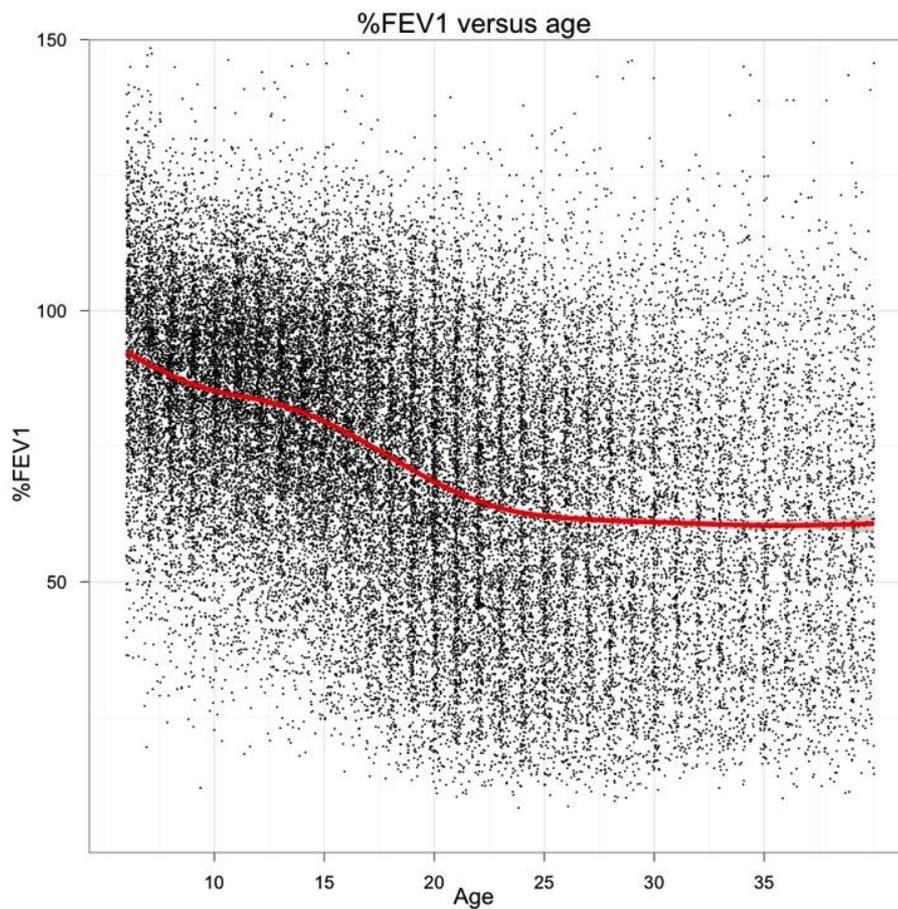
The %FEV<sub>1</sub> and *P. aeruginosa* analyses are now considered separately in more detail.

### ***Lung function (%FEV<sub>1</sub>)***

Mean cross-sectional %FEV<sub>1</sub> falls linearly with age, and then plateaus out between age 20 and 25 (Figure 45). This was modelled as a linear trend in time in the <18 age group, and as a split line function in the >18 age group (see Appendix 4, Figure 91 for further exploratory analysis plots illustrating cross-sectional trajectories stratified by baseline covariates).

**Figure 45: %FEV<sub>1</sub> versus age from 5 to 40**

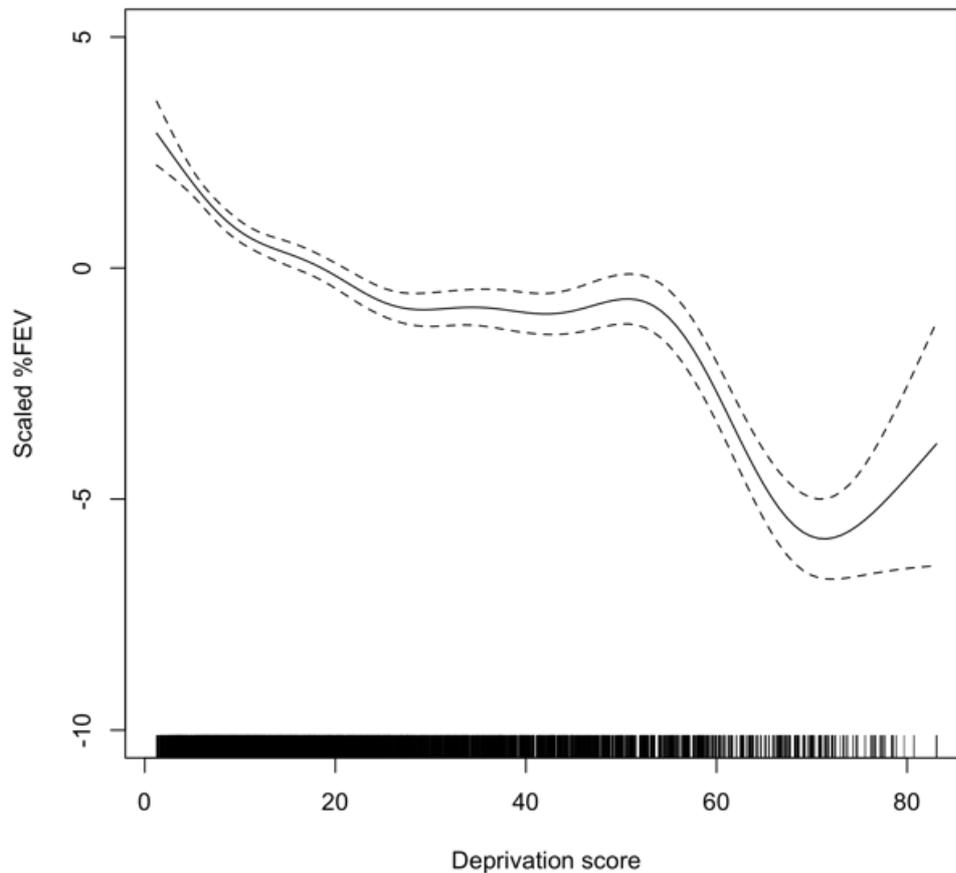
*Mean smoother in red.*



Exploratory analysis using a generalized additive model (GAM) suggested a dose-response relationship between deprivation score and level of %FEV<sub>1</sub> (Figure 46).

**Figure 46: GAMs showing the shape of the relationship between %FEV<sub>1</sub> and deprivation score**

*%FEV<sub>1</sub> decreases with increasing deprivation, and there is a dose-response relationship.*



In the final model for %FEV<sub>1</sub>, in the paediatric age range there was a difference of -4.1 (-5 to -3.1) comparing children in the most deprived quintile to the least, apparent as soon as %FEV<sub>1</sub> can be measured around age five years, but there was no evidence of an increased rate of decline of %FEV<sub>1</sub> (Table 8, Figure 47). The estimates were consistent with a monotonic dose-response relationship between deprivation and %FEV<sub>1</sub> (Figure 46). In the paediatric age group, higher %FEV<sub>1</sub> was associated with male sex, screened patients, heterozygote delta F508 status, white patients, no CFRD, no *P. aeruginosa* colonisation, and higher BMI (Table 8, Figure 47). Adding BMI SD score to the model reduced the %FEV<sub>1</sub> deprivation gap to -3.5 (95% CI -5.2 to -1.8). Further adjusting for *B. cepacia* status and care-centre did not change the deprivation effect on %FEV<sub>1</sub> (see robustness test section).

**Table 8: Final regression models for %FEV<sub>1</sub> in <18 age group**

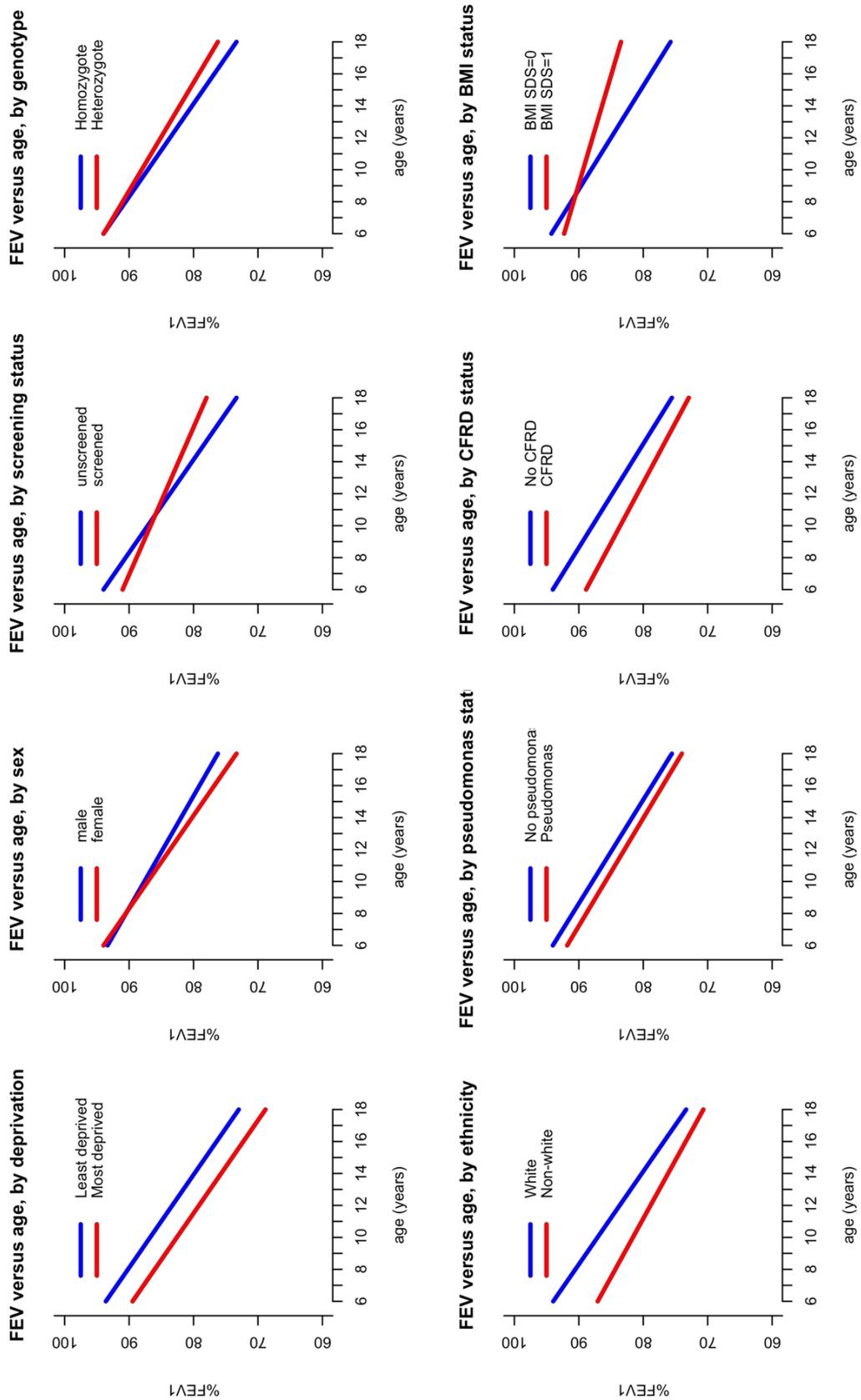
	<i>final</i>	<i>Add time varying</i>	<i>Add BMI</i>
Constant	95.688*** (1.276)	95.568*** (1.557)	95.794*** (1.529)
Age-5	-1.719*** (0.085)	-1.539*** (0.136)	-1.541*** (0.133)
Number of F508 alleles: 0/2	-0.142 (1.402)	-0.258 (1.402)	-0.980 (1.411)
Number of F508 alleles: 1/2	-0.258 (0.829)	-0.346 (0.828)	-0.062 (0.835)
Male	-0.924 (0.768)	-0.909 (0.764)	-1.355 (0.770)
nonwhite	-7.256*** (2.149)	-7.114*** (2.134)	-8.680*** (2.147)
Screened	-3.599*** (1.064)	-3.584*** (1.057)	-3.435** (1.065)
Age-5 x nallele: 0/2	0.016 (0.161)	-0.003 (0.161)	0.135 (0.157)
Age-5x nallele: 1/2	0.242* (0.100)	0.233* (0.100)	0.180 (0.097)
Age-5 x sex: Male/Female	0.293** (0.094)	0.274** (0.093)	0.352*** (0.090)
Age-5 x nonwhite	0.353 (0.266)	0.361 (0.265)	0.742** (0.257)
Age-5 x screened	0.634*** (0.131)	0.621*** (0.130)	0.585*** (0.126)
Deprivation score	-0.071*** (0.016)	-0.070*** (0.016)	-0.061*** (0.015)
<i>Pseudomonas</i> colonisation		-2.282*** (0.609)	-2.548*** (0.602)
CFRD		-5.373** (1.980)	-7.269*** (1.945)
Pancreatic insufficiency		0.380 (0.954)	-0.094 (0.945)
Age-5 x <i>Pseudomonas</i> colonisation		0.056 (0.073)	0.115 (0.072)
Age-5 x CFRD		0.212 (0.198)	0.396* (0.194)
Age-5 x Pancreatic insufficiency		-0.129 (0.114)	-0.024 (0.112)
BMI Z score			-2.784*** (0.107)
Age-5 x BMI Z score			0.806*** (0.027)
Log-likelihood	-80509.181	-80458.674	-79894.252
Deviance	161018.362	160917.348	159788.503
AIC	161100.362	161011.348	159886.503
BIC	161424.362	161382.763	160273.668
N	19979	19979	19957
Groups	4445	4445	4443

\* P < 0.05, \*\* P < 0.01, \*\*\* P < 0.001

Standard errors in parentheses, birthyear coefficients not shown

The deprivation effect is multiplied by 58 to generate the contrast between the mid point of the least and most deprived quintile

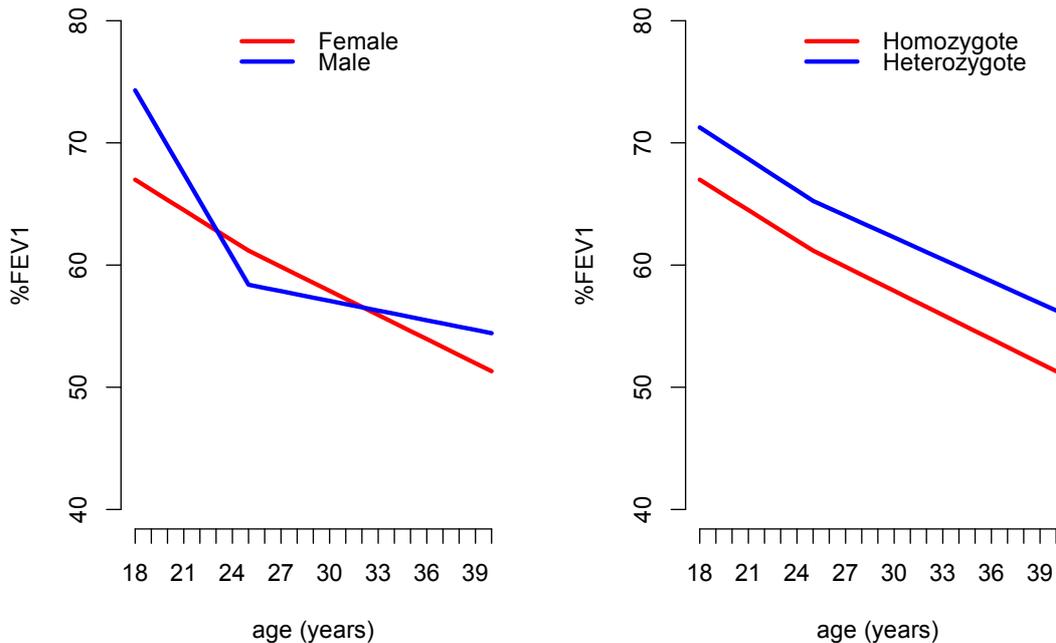
**Figure 47: %FEV<sub>1</sub> trajectories, illustrating the effect of deprivation, sex, screening, genotype, ethnicity, *Pseudomonas*, CFRD and BMI status**



There was no significant association between %FEV<sub>1</sub> and social deprivation in the adult age range -1.6 (95% CI -4.4 to 1.25) (Appendix 4, Table 26, Figure 91) though there were significant sex, and genotype effects, as illustrated in Figure 48: male patients initially have higher lung function, but this effect is attenuated at older ages, due to an increased rate of decline in men before the age of 25, whereas heterozygotes for delta F508 status consistently have higher %FEV<sub>1</sub>.

**Figure 48: %FEV<sub>1</sub> versus age in the >18 group, illustrating sex and genotype effect**

*Trajectories plotted at reference values for other covariates in the regression model.*

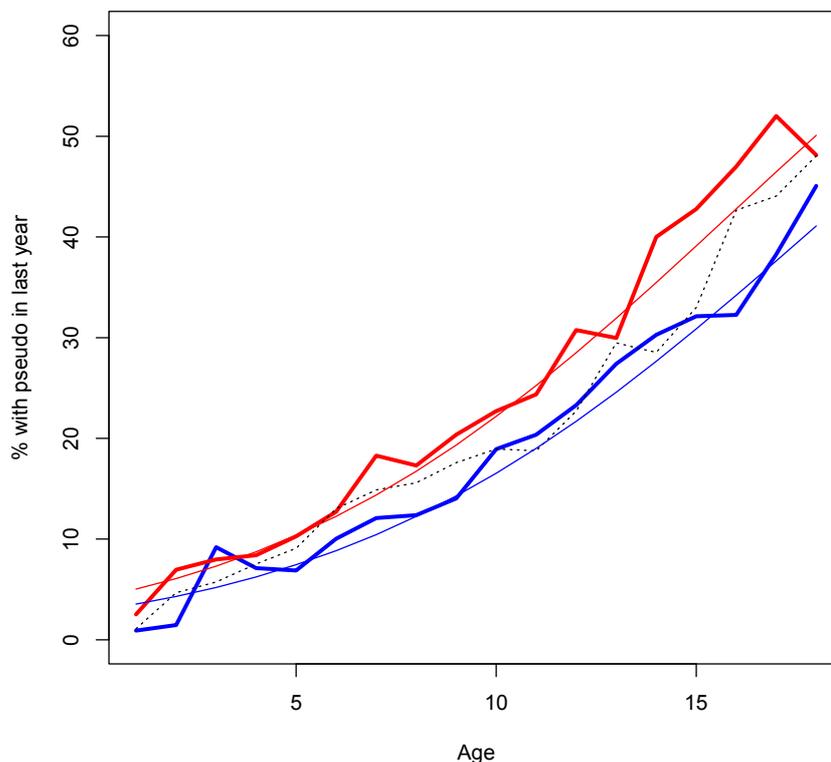


### ***Pseudomonas* colonisation**

The cross-sectional proportion of people with chronic *P. aeruginosa* increased steadily with age to around 60% by age 20, and was more common in the most deprived quintile, with an OR of 1.9 (95% CI 1.3 to 2.7) in the adjusted paediatric analysis for the most deprived quintile (Figure 49). Increased likelihood of *P. aeruginosa* colonisation was associated with female sex, homozygote delta F508 status, CFRD, PI, and lower %FEV<sub>1</sub>, but adjusting for these factors did not substantially alter the deprivation effect. The estimates were consistent with a monotonic dose-response relationship between deprivation and risk of *P. aeruginosa* (Figure 50). The deprivation effect was similar in the adult age range with an OR of 1.78 (95% CI 1.26 to 2.51) comparing the most to the least deprived quintile (Appendix 4, Table 27, Table 28).

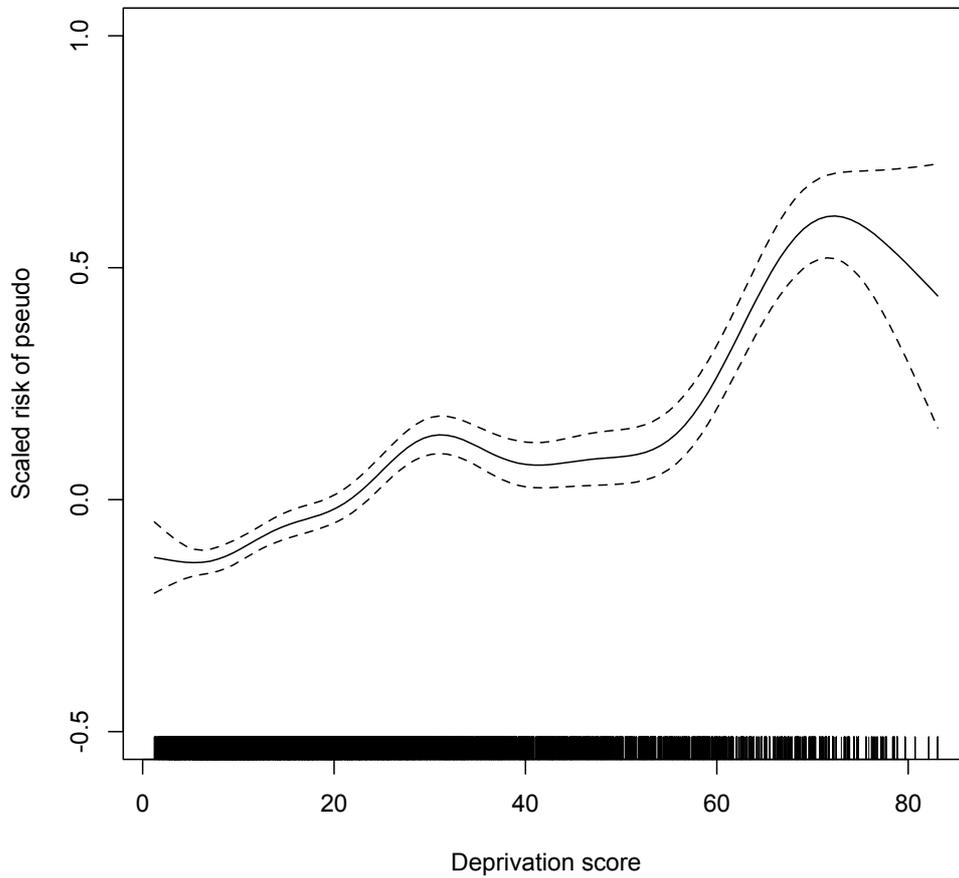
**Figure 49: Cross sectional and modelled longitudinal *Pseudomonas* prevalence by deprivation**

*Thick lines show the cross-sectional prevalence, and thin lines show the modelled longitudinal population average *Pseudomonas* prevalence for the most deprived (red) and least deprived (blue) quintiles*



**Figure 50: GAMs showing the shape of the relationship between risk of *Pseudomonas* colonisation and deprivation score**

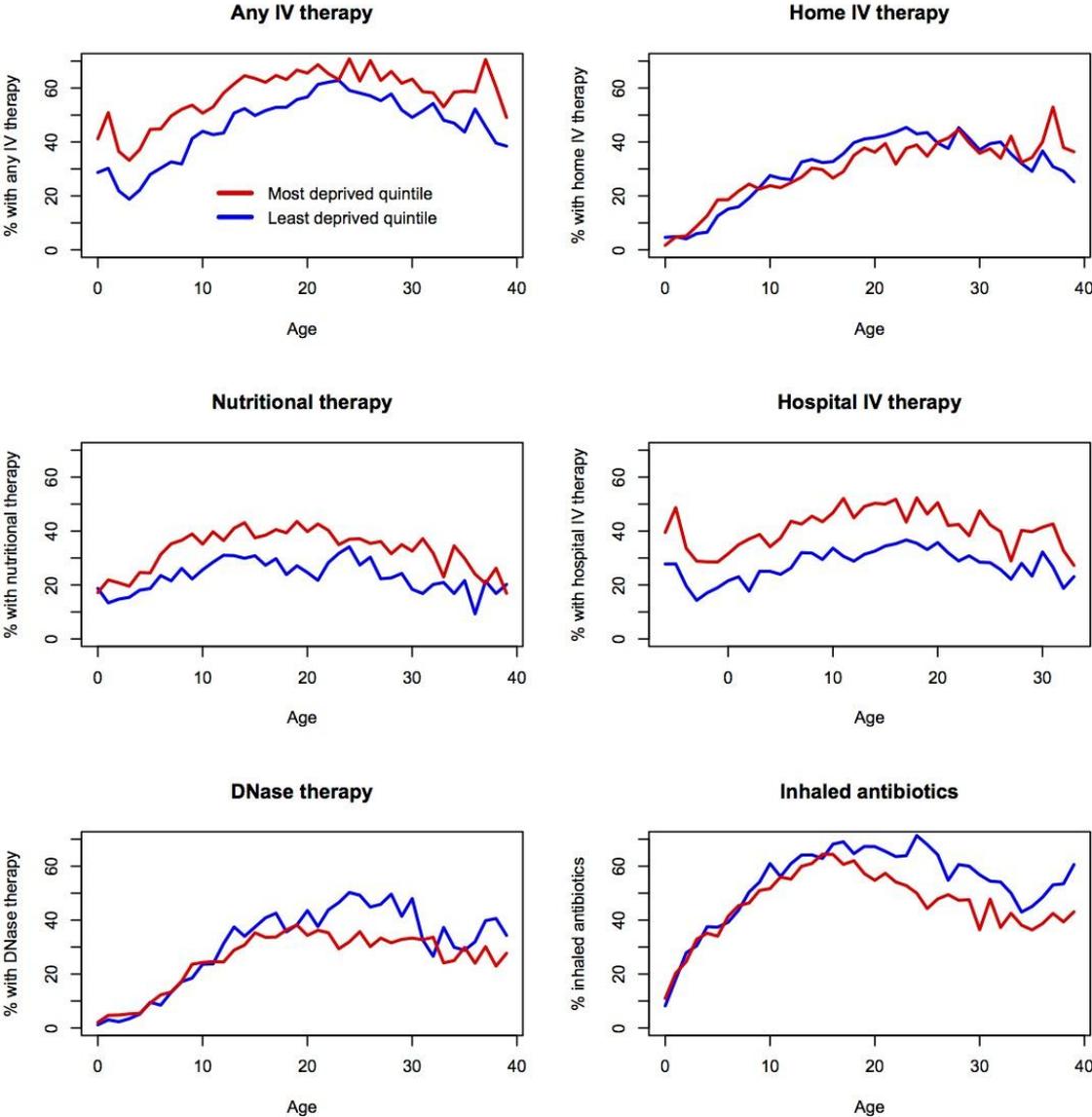
*Risk of Pseudomonas colonisation increases with increasing deprivation, and there is a dose-response relationship.*



### **Use of therapies**

First I provide an overall cross-sectional summary for the use of therapy outcomes (IV therapy, nutritional, and inhaled therapies). People with CF from the most deprived areas in the UK are more likely to receive any IV therapy and supplemental feeding, compared to the least deprived areas, after adjustment for disease severity, but they are less likely to receive IV therapy at home, DNase and inhaled antibiotic therapy (Figure 51).

**Figure 51: Use of therapies: any IV antibiotic therapy, home IV antibiotic therapy, hospital IV antibiotic therapy, supplemental feeding, DNase, and inhaled antibiotics, by age comparing extremes of deprivation quintile**



**IV therapy**

The final regression model parameters for use of any IV therapy are shown in Table 9. The use of any IV therapy, after adjusting for disease severity, was over twice as common in the most deprived children aged over five (OR 2.52, 95% CI 1.92 to 3.17, Table 9), and this trend continued in adults (OR 1.89, 95% CI 1.51 to 2.38, Appendix 4, Table 29).

**Table 9: Final regression models for any IV therapy in 5 to <18 age group**

	<i>severity</i>	<i>final</i>	<i>final_bmi</i>
Constant	-1.450*** (0.196)	-1.800*** (0.202)	-1.799*** (0.202)
Age-5	0.341*** (0.036)	0.340*** (0.036)	0.339*** (0.036)
(Age-5)^2	-0.016*** (0.002)	-0.016*** (0.002)	-0.016*** (0.002)
Number of F508 alleles: 0/2	-0.499*** (0.127)	-0.533*** (0.127)	-0.531*** (0.127)
Number of F508 alleles: 1/2	-0.366*** (0.079)	-0.388*** (0.079)	-0.390*** (0.079)
Male	-0.365*** (0.074)	-0.374*** (0.073)	-0.373*** (0.073)
%FEV <sub>1</sub>	-0.036*** (0.002)	-0.036*** (0.002)	-0.035*** (0.002)
<i>Pseudomonas</i> colonisation	1.729*** (0.065)	1.720*** (0.065)	1.717*** (0.065)
Deprivation		0.016*** (0.002)	0.016*** (0.002)
BMI SD score			-0.056 (0.030)
Log-likelihood	-9469.776	-9447.085	-9436.755
Deviance	18939.553	18894.170	18873.509
AIC	19009.553	18966.170	18947.509
BIC	19282.462	19246.877	19235.974
N	17987	17987	17968
Groups	4321	4321	4319

\*P < 0.05, \*\* P < 0.01, \*\*\* P < 0.001

NB parameters represent log-odds

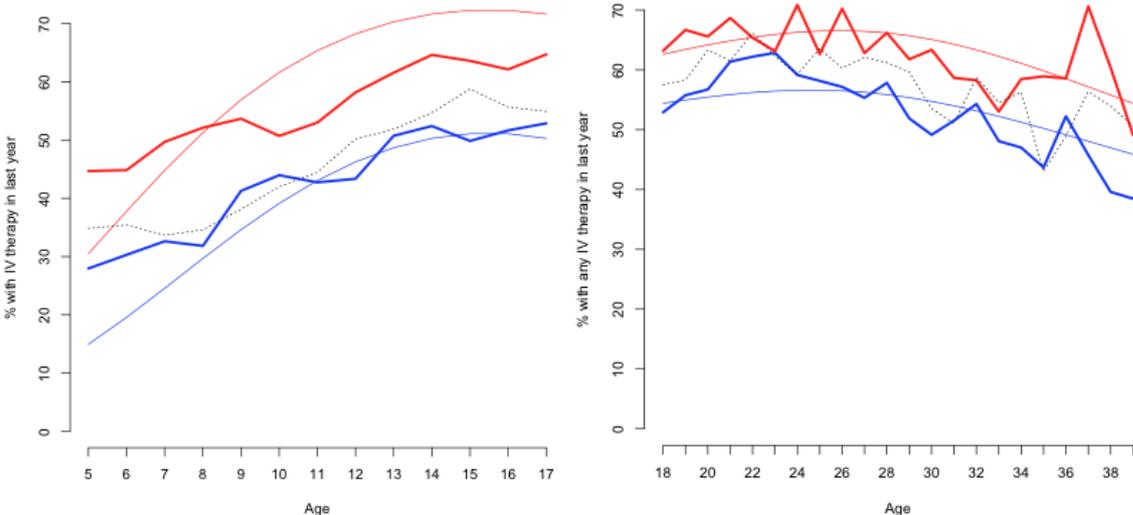
Standard errors in parentheses, birthyear coefficients not shown

The deprivation effect is multiplied by 58 to generate the contrast between the mid point of the least and most deprived quintile

The deprivation effect on use of any IV therapy is visualised in Figure 52. The estimates were consistent with a monotonic dose-response relationship between SES and use of any IV therapy (Figure 53), and further adjusting for care-centre did not change this effect (Appendix 4, Table 30). Use of any IV therapy was associated with female sex, homozygote delta F508 status, low %FEV<sub>1</sub> and *Pseudomonas* colonisation (Table 9).

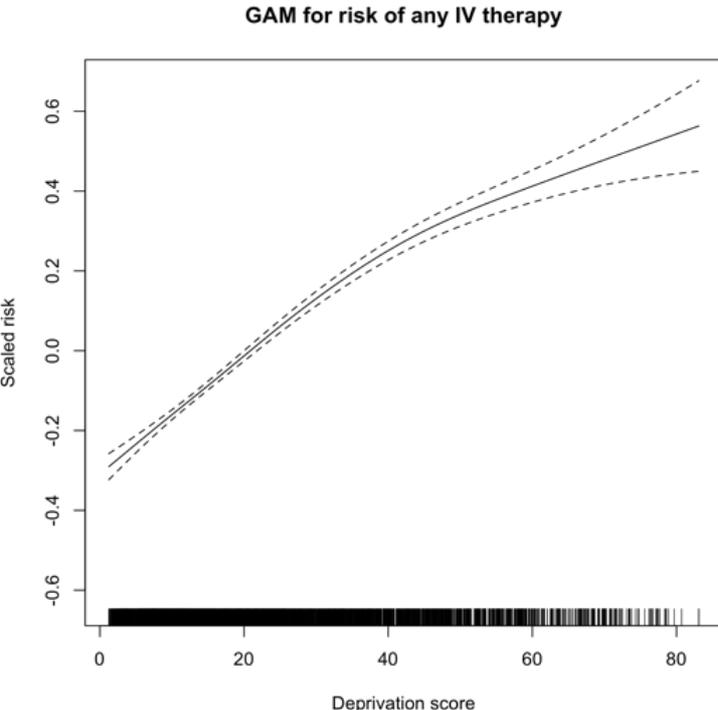
**Figure 52: Cross sectional and modelled longitudinal percentage of people using any IV therapy in the preceding year by deprivation in the <18 and >18 age group**

*Thick lines show the cross-sectional prevalence, and thin lines show the modelled longitudinal population averages for the most deprived (red) and least deprived (blue) quintiles*



**Figure 53: GAMs showing the shape of the relationship between log-odds of any IV therapy and deprivation score.**

*The likelihood of receiving any IV therapy increases with increasing deprivation, and there is a dose-response relationship.*

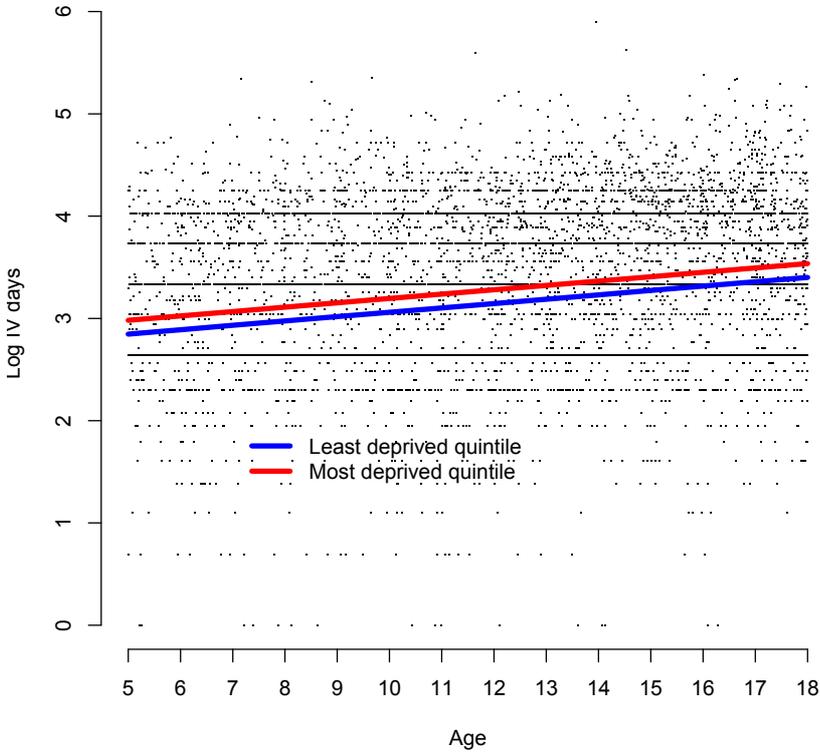


When hospital and home IV therapy were analysed separately, it became clear that the higher prevalence of any IV therapy observed in the most deprived quintile was almost entirely due to delivery of IV therapy in hospital, rather than home as indicated in Figure 51. For home IV therapy there was evidence of an interaction between deprivation and time, such that the chance of home IV therapy became more common in the least deprived quintile with increasing age (Appendix 4, Table 31, Table 32).

Conditional on IV therapy use, I modelled the log of the total number of IV days per year as a function of time and other covariates. The deprivation effect is demonstrated in Figure 54, which shows that people from the most deprived quintile, compared to the least, have more IV therapy, conditional on having any. After adjusting for disease severity, people in the most deprived quintile had 15.9% (95% CI 8.2 to 24) more IV therapy days in the paediatric age range, and 10.6% (95% 2.5 to 19.2) more IV days in the adult age range per year. Use of more IV therapy days was associated with female sex, homozygote delta F508 status, low %FEV<sub>1</sub> and *Pseudomonas* colonisation (see final regression tables in Appendix 4, Table 33, Table 34).

**Figure 54: Scatterplot of log IV days (home and hospital) versus age, with fitted longitudinal trajectory by deprivation quintile**

*Entry to the analysis is conditional on having one or more IV days*

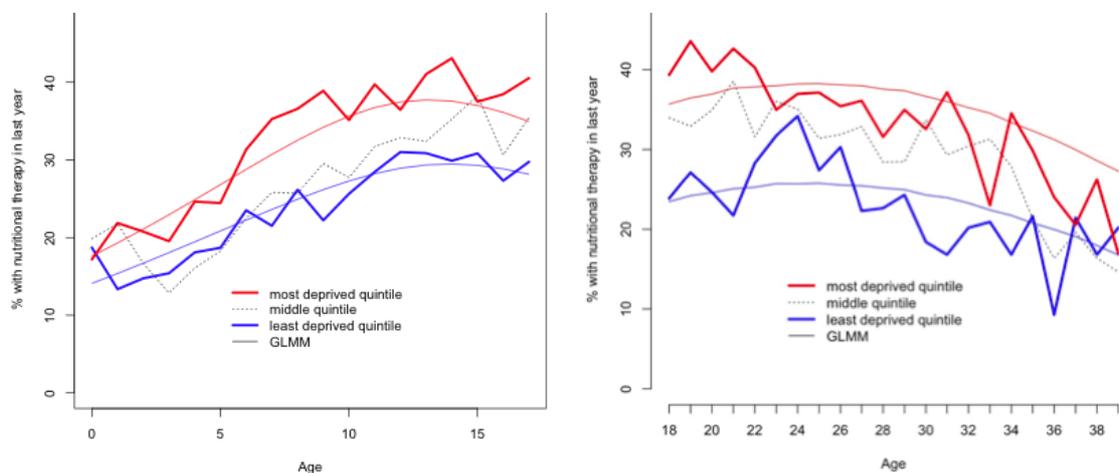


### ***Nutritional therapy***

Prevalence of any supplemental feeding therapy in the previous year was more common in the most deprived quintile, compared to the least, across the entire age range from 0 to age 40 (OR 1.78, 95% CI 1.42 to 2.2, adjusted for baseline variables, *P. aeruginosa* status and BMI, in the 5 to 18 age group) (Figure 55). The estimates were consistent with a monotonic dose-response relationship between SES and use of any nutritional therapy. Use of any nutritional therapy in the preceding year was associated with homozygote delta F508 status, low BMI and *Pseudomonas* colonisation (regression models for nutritional therapy in Appendix 4, Table 35, Table 36).

**Figure 55: Cross sectional and modelled longitudinal percentage of people using any nutritional therapy in the preceding year by deprivation in the <18 age group and <18 age group**

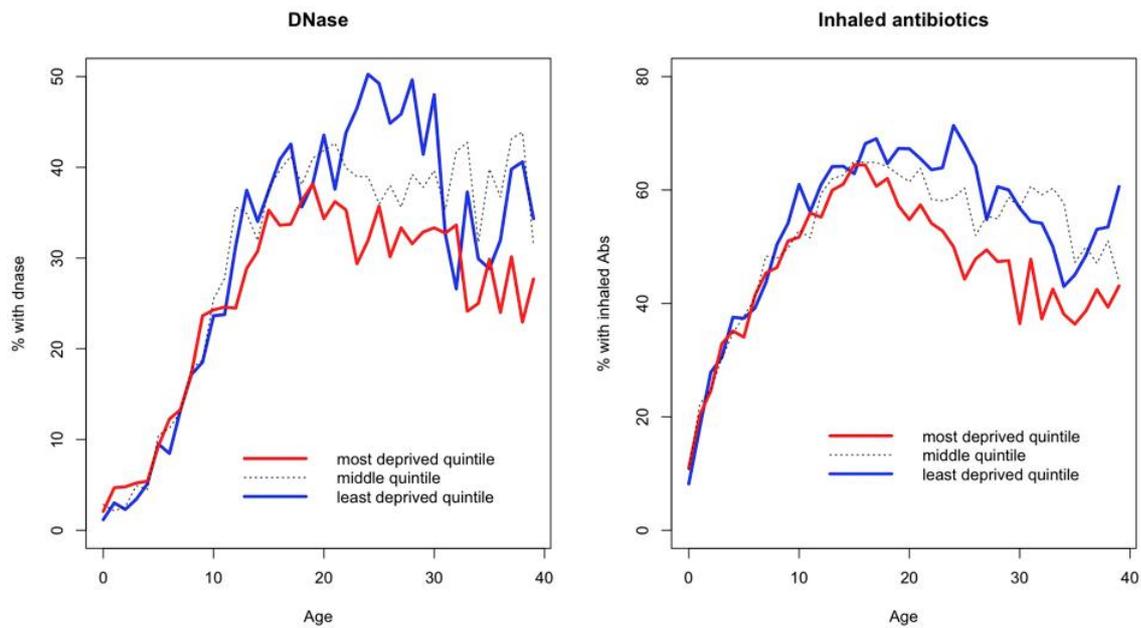
*Thick lines show the cross-sectional prevalence, and thin lines show the modelled longitudinal population averages for the most deprived (red) and least deprived (blue) quintiles*



### *Inhaled therapies*

There was no significant association between DNase use and deprivation in the paediatric age range before adjusting for disease severity. After adjustment for disease severity, treatment was less likely in the most deprived quintile, in both children (OR 0.40 95% CI 0.21 to 0.72) and adults (OR 0.37 95% CI 0.26 to 0.52), though the association with deprivation was stronger in adults. A similar pattern was found for inhaled antibiotic therapy (Table 5, Figure 56, final regression tables in appendix Appendix 4, Table 37, Table 38, Table 39, Table 40).

**Figure 56: Cross sectional percentage of people using any DNase therapy, and inhaled antibiotic, by deprivation quintile**



## Robustness checks

Illustrative residual diagnostics for the models for %FEV<sub>1</sub> and any IV therapy are in the appendix (Appendix 4, Figure 92, Figure 93). In addition, I undertook a range of additional analyses as robustness checks, including a comparison of the excluded and included population (Table 10). The excluded population is shifted towards older birth cohorts, reflecting improved collection of postcodes over time.

**Table 10: Comparison of characteristics of eligible population versus those not meeting the inclusion criteria**

	<i>Excluded</i>	<i>Included</i>
Total (%)	1198	8055
Female (%)	587 (49)	3764 (46.7)
Male (%)	611 (51)	4291 (53.3)
No. delta 508: 2 (%)	621 (51.8)	4159 (51.6)
No. delta 508: 1 (%)	356 (29.7)	2862 (35.5)
No. delta 508: 0 (%)	221 (18.4)	1034 (12.8)
Birth cohort >1957-01-01 (%)	77 (6.4)	261 (3.2)
>1967-01-01 (%)	235 (19.6)	835 (10.4)
>1977-01-01 (%)	415 (34.6)	1904 (23.6)
>1987-01-01 (%)	300 (25)	2528 (31.4)
>1997-01-01 (%)	143 (11.9)	2022 (25.1)
>2007-01-01 (%)	28 (2.3)	505 (6.3)
White (%)	1155 (96.4)	7748 (96.2)
Died	490 (40.9)	435 (5.4)
%FEV <sub>1</sub> at age 6 (median and interquartile range)	94.29 (80.26 - 103.70)	92.93 (80.26 - 103.70)
% with <i>Pseudomonas</i> at age 6 (95% CI)	8.8 (4.5 - 16.5)	10.9 (9.5 - 12.6)

The robustness checks were applied to the final model for %FEV<sub>1</sub> (Table 11), and the final model for any IV therapy (Appendix 4, Table 30).

### *Changing deprivation scores*

Over the study period 18% of eligible individuals had more than one postcode recorded. As a robustness check, I repeated the analysis for %FEV<sub>1</sub>, treating SES as a time-varying covariate, but this did not substantially alter the result (Table 11).

### *Adjustment for clustering by CF centre*

Differences between centres may mediate some of the effects of SES on outcomes, and explain some of the differences in treatments received. In order to explore this I

replicated the final models for %FEV<sub>1</sub>, and for any IV therapy, adding in care centre as a fixed effect. This made no difference to the deprivation effect (Table 11).

***Excluding data pre-2000 or only including data >2005***

Excluding the data pre-2000, when recruitment to the cohort was increasing over time made no difference to the deprivation effect. Including data only after the move to Port CF also did not have a substantive effect on the conclusions (Table 11).

***Adjusting for *B. cepacia* status***

Of individuals in the %FEV<sub>1</sub> analysis for the <18 group, 3.5% (156/4445) had *B. cepacia*. Addition of this variable to the model for %FEV<sub>1</sub> made no difference to the deprivation effect (Table 11).

***Restricting the analysis to English data only***

This did not have any substantive effect on the conclusions (Table 11).

***Re-fitting models with deprivation treated as a five-level factor***

This did not have any substantive effect on the conclusions for any of the models (Table 12).

**Table 11: Additional models fitted as robustness tests, based upon the final FEV<sub>1</sub> model**

	final	IMD as time varying	Final z score	Final + care center	Data <2000 excluded	Cepacia added	English data	Data>2005
Constant	95.688*** (1.276)	95.726*** (1.274)	94.137*** (1.225)	89.83*** (1.88)	95.609*** (1.303)	95.641*** (1.274)	96.176*** (1.408)	96.133*** (1.818)
Age-5	-1.719*** (0.085)	-1.719*** (0.085)	-1.719*** (0.085)	-1.69*** (0.09)	-1.722*** (0.089)	-1.705*** (0.085)	-1.712*** (0.094)	-1.771*** (0.142)
Number of F508 alleles (nallele): 0/2	-0.142 (1.402)	-0.135 (1.402)	-0.143 (1.402)	-0.54 (1.38)	-0.526 (1.452)	-0.140 (1.401)	-0.787 (1.587)	1.868 (2.041)
Number of F508 alleles: 1/2	-0.258 (0.829)	-0.260 (0.829)	-0.259 (0.829)	-0.19 (0.81)	-0.258 (0.863)	-0.237 (0.828)	-0.694 (0.938)	-0.601 (1.220)
Male	-0.924 (0.768)	-0.921 (0.768)	-0.925 (0.768)	-1.11 (0.75)	-1.073 (0.800)	-0.919 (0.767)	-1.065 (0.863)	-0.257 (1.136)
nonwhite	-7.256*** (2.149)	-7.248*** (2.148)	-7.252*** (2.149)	-7.36*** (2.13)	-6.635** (2.203)	-7.176*** (2.146)	-7.323** (2.234)	-7.632** (2.875)
Screened	-3.599*** (1.064)	-3.581*** (1.064)	-3.597*** (1.064)	-2.63* (1.15)	-3.809*** (1.101)	-3.601*** (1.062)	-2.640* (1.320)	-2.461 (1.536)
Deprivation score	-0.071*** (0.016)			-0.07*** (0.02)	-0.075*** (0.016)	-0.071*** (0.016)	-0.082*** (0.017)	-0.080*** (0.019)
Age-5 x nallele: 0/2	0.016 (0.161)	0.017 (0.161)	0.017 (0.161)	0.09 (0.16)	0.065 (0.166)	0.014 (0.160)	-0.023 (0.181)	-0.128 (0.243)
Age-5x nallele: 1/2	0.242* (0.100)	0.242* (0.100)	0.242* (0.100)	0.25* (0.10)	0.228* (0.104)	0.242* (0.100)	0.285* (0.113)	0.309* (0.154)
Age-5 x sex: Male/Female	0.293** (0.094)	0.292** (0.094)	0.293** (0.094)	0.29** (0.09)	0.302** (0.098)	0.292** (0.094)	0.294** (0.105)	0.233 (0.145)
Age-5 x non white	0.353 (0.266)	0.354 (0.266)	0.353 (0.266)	0.31 (0.27)	0.285 (0.273)	0.343 (0.266)	0.444 (0.276)	0.517 (0.376)
Age-5 x screened	0.634*** (0.131)	0.631*** (0.131)	0.634*** (0.131)	0.53*** (0.13)	0.667*** (0.136)	0.637*** (0.131)	0.519** (0.160)	0.604** (0.198)
Deprivation score (time varying)		-0.073*** (0.016)						
Deprivation Z score			-1.128*** (0.251)					
Cepacia						-2.302 (2.415)		
Age-5 x cepacia						-0.194 (0.271)		
Log-likelihood	-80509.181	-80508.314	-80509.171	-80316.11	-74959.478	-80499.082	-66386.620	-
Deviance	161018.362	161016.629	161018.341	160632.21	149918.955	160998.164	132773.239	71820.788
AIC	161100.362	161098.629	161100.341	160976.21	149994.955	161084.164	132855.239	71886.788
BIC	161424.362	161422.628	161424.341	162335.43	150292.483	161423.969	133171.196	72120.466
N	19979	19979	19979	19979	18577	19979	16420	8788
Groups	4445	4445	4445	4445	4334	4445	3807	3318

\*P < 0.05, \*\* P < 0.01, \*\*\* P < 0.001

Standard errors in parentheses, birthyear coefficients not shown

The deprivation effect is multiplied by 58 to generate the contrast between the mid point of the least and most deprived quintile

**Table 12. Regression coefficients from explanatory models fitting deprivation as a five-level factor**

	<b>Weight Z score</b>	<b>SE</b>	<b>Height Z score</b>	<b>SE</b>	<b>BMI Z score</b>	<b>SE</b>		
quintile 2	-0.020	0.050	-0.030	0.050	-0.010	0.040		
quintile 3	-0.030	0.050	-0.070	0.050	0.020	0.040		
quintile 4	-0.050	0.040	-0.110	0.040	0.030	0.040		
quintile 5	-0.230	0.040	-0.250	0.040	-0.120	0.040		
	<b>%FEV<sub>1</sub></b>	<b>SE</b>	<b><i>Pseudomonas</i> colonisation</b>	<b>SE</b>				
quintile 2	-0.900	0.850	0.140	0.170				
quintile 3	-1.560	0.860	0.140	0.170				
quintile 4	-1.330	0.830	0.190	0.160				
quintile 5	-3.370	0.840	0.440	0.160				
	<b>Any IV therapy</b>		<b>Nutritional therapy</b>	<b>SE</b>	<b>DNase</b>	<b>SE</b>	<b>Inhaled antibiotics</b>	<b>SE</b>
quintile 2	0.080	0.120	0.050	0.110	-0.210	0.160	-0.040	0.140
quintile 3	0.190	0.120	0.190	0.110	-0.050	0.160	-0.220	0.140
quintile 4	0.430	0.110	0.190	0.110	-0.450	0.160	-0.520	0.140
quintile 5	0.800	0.120	0.610	0.110	-0.570	0.160	-0.820	0.140

## Discussion

Children with CF from the most disadvantaged areas in the UK have significantly lower weight, height and BMI in the first years of life after diagnosis, are more likely to have chronic *P. aeruginosa* infection, and a significantly lower %FEV<sub>1</sub> when it can first be measured (i.e. after the age of five), but these inequalities do not increase over time.

There is evidence of positive discrimination, or ‘pro-poor’ bias, in the provision of some key treatments, on the basis of socio-economic circumstances. I show that in the NHS, children with CF from the most disadvantaged areas are about twice as likely to receive IV antibiotic therapy (specifically in hospital) and nutritional support, after adjustment for disease severity. I also show some apparent ‘pro-rich’ bias in two other therapies – DNase and inhaled antibiotics – in that patients from the most affluent areas are significantly more likely to receive these treatments after adjustment for disease severity.

Key strengths of this study include the population-wide coverage of the UK CF registry, the high quality of the data and the longitudinal analysis. There are limitations: First, it relies on retrospective, routinely collected data and use of a standard measure of deprivation of area of residence, the IMD. Each small area contains about 1500 people, and in this respect the IMD allows much finer resolution than the previous US analyses (O'Connor et al., 2003, Schechter et al., 2009, Schechter et al., 2011) that have used ZIP code linked income data, since each ZIP code contains about 30,000 people (Krieger et al., 2002). There is always the possibility of ecological fallacy, but this is unlikely, given that similar associations have been found in the US studies that use both area and individual measures of SES. Second, I had valid postcodes on only 90% of the sample, although my sample size was large, with no pronounced gradient in the proportion of patients by deprivation quintile. The excluded population is shifted towards older birth cohorts, reflecting improved collection of postcodes over time, but I do not believe that this has significantly biased the associations observed (Table 10). Third, there is a strong cohort effect in CF and, with datasets of this type, age and cohort effects confound one another, and cannot be completely separated (Taylor-Robinson et al., 2012a). I have adjusted for both in our analysis, to estimate the adjusted effect on deprivation.

This study quantifies the longitudinal weight trajectory of the UK CF population, and the effects of area deprivation, sex and genotype. In the first three years of life there is a period of improvement in weight SD score, presumably secondary to diagnosis and subsequent pancreatic enzyme supplementation with close attention to nutrition. Similar patterns have been observed in cross-sectional analyses of US registry data (Lai et al., 1999). Our study identified significant differences in sex-adjusted WAZ, with males being heavier from the outset. Other studies have shown differences in weight trajectories by sex in CF (Zemel et al., 2000), but when British reference standards are used, this sex difference appears greater, so this finding may partly represent a standardisation artefact (Lai et al., 1999). In addition genotype also had an effect, with homozygotes for delta F508 experiencing greater weight gain in the first three years, and greater weight loss subsequently. This probably relates to a greater diagnosis and treatment effect early on in life, followed by accelerated weight loss subsequently as a result of pancreatic insufficiency.

Overall, the UK CF population is significantly underweight, and shorter compared to the UK reference population, by about a third of an SD score. Deprivation approximately doubles this effect, lowering the SD score by another third. It is not clear how much of the effect of SES on growth outcomes is specific to CF, and how much is a reflection of SES effects in the general population. Comparable data in contemporary representative cohorts in the UK is lacking, but the age related changes in growth in the general population are characterised by increasing obesity in childhood from the age of 4 onwards, with higher BMI in the more deprived populations (Howe et al., 2011, Howe et al., 2012), and this is opposite to the patterns seen in our study. In a large cohort, birth weight was lower in more deprived populations in the UK, by about 0.2 of an SD score, when comparing the most deprived Townsend score quartile to the least (Kinra et al., 2005). This difference is smaller than the projected weight difference at intercept in our study (-0.54), but this difference could be influenced by differences in the size of the subgroups compared (quartiles in the former and quintiles in ours). Another recent study found no difference in birth weight by SES measured using parental occupational class (NS-SEC) (Wijlaars et al., 2011). I speculate that having CF is likely to amplify the effects of SES on nutritional status at birth, and in the first few years of life.

The inequality in weight is greatest at around the time of diagnosis, and becomes narrower over the first three years of life. This is an important finding, since a widening of inequalities over time is often the norm (Ben-Shlomo and Kuh, 2002, Whitehead and Dahlgren, 2007, CSDH, 2008). These findings suggest that extending the period of differential weight gain for as long as possible may reduce inequalities, further supporting neonatal screening programmes to facilitate early diagnosis and therapy (Southern et al., 2009). We can speculate that by extending this period of catch-up for as long as possible by early diagnosis (i.e. screening) we may see an attenuation of the deprivation effect over time. In this study, there was no difference in the age at diagnosis by deprivation, but screening was associated with improved weight, height, and lung function in children. Furthermore, our finding that the prevalence of supplemental feeding therapy was higher, after adjusting for disease severity, in the most disadvantaged patients suggests that NHS professionals are actively engaged in trying to boost the nutrition of poorer patients in particular, recognising their health disadvantage.

The social gradient in lung function, evident as soon as it can be routinely measured at age five, points to the crucial role of environmental and health care factors operating in the early years of life to produce inequalities. It further reinforces the need for early diagnosis and action to prevent adverse consequences for children with CF living in disadvantaged circumstances. In Schechter and colleagues' cross-sectional study of US data inequalities in %FEV<sub>1</sub> by Medicaid status widened slightly from 5 to 20 years of age (Schechter et al., 2001). The magnitude of the inequalities in lung function at age five found in Schechter's study was larger (around 9% difference) than in our UK study (4%), as was the magnitude of inequalities in lung function found in O'Connor and colleagues' US study, which found a difference of 5.5% between most and least deprived quintiles (O'Connor et al., 2003). Methodological differences between the studies, however, make a direct comparison between UK and US findings on the *magnitude* of the deprivation gap in lung function inappropriate. This study is the first to examine the relationship between deprivation and %FEV<sub>1</sub> in a population level adult cohort. I did not find a significant association, despite the higher prevalence of *P. aeruginosa*. I speculate that this finding may relate to the complication of progressive drop-out in older

patients, and the relative insensitivity of %FEV<sub>1</sub> as an outcome measure in adults (Taylor-Robinson et al., 2012a).

The increased prevalence of chronic *P. aeruginosa* infection in patients from more deprived areas, after adjusting for %FEV<sub>1</sub>, is an important new finding in a population-level cohort. In Schechter and colleagues study (Schechter et al., 2001), Medicaid patients were more likely than non-Medicaid patients to have *P. aeruginosa*, but when adjusted for %FEV<sub>1</sub> there was no statistical difference, and a more recent US cohort study did not demonstrate an association (Rosenfeld et al., 2012a). Previously identified risk factors for *P. aeruginosa* acquisition, which is associated with worse lung function, include female sex and genotype (both significant in this study), and exposure to other patients with *P. aeruginosa* colonisation (Schechter, 2011). My finding that more deprived groups are more likely to receive IV therapies in hospital may result in more deprived patients having greater exposure to other patients with chronic *P. aeruginosa*.

Marked socio-economic differences were observed in the reported use of key CF therapies in two contrasting ways. First, children from the most deprived quintile were around twice as likely to receive hospital IV antibiotic therapy, and nutritional support, after adjustment for disease severity. I can speculate, from my knowledge of UK CF services, that clinicians in the NHS are more likely to bring children from more deprived areas into hospital for IV therapies because of concerns about the difficulties in delivering treatments in their home environment. Conversely, children living in more affluent circumstances may receive IV therapy at home because of judgements about the adequacy of support and adherence to therapy in the home and/or because of the families' wish to avoid disruption to schooling and family life. This equitable model of care, with positive discrimination for socially disadvantaged children and adults with CF, is an uncommon finding in health systems, where often access, particularly to secondary care for adults, declines with increasing deprivation, after adjusting for differential need (Hanratty et al., 2007b, Stirbu et al., 2011). While several studies have found use of health services in general by level of deprivation, adjusted for need, to be more equal in relation to children than adults (Groholt et al., 2003), I have found evidence in CF children that goes even further with a pro-poor bias in the NHS for specific therapies. Coupled with my findings of inequalities in outcomes by deprivation which do not widen over time, I speculate that the treatment

decisions being made by clinicians may mitigate some effects of social disadvantage. This provides encouragement that there are interventions that health services can make to reduce the adverse effects of deprivation on chronic conditions such as CF. In the US, using zip-code linked income of an area as the socio-economic indicator, there was no gradient in IV therapy use in children <12, but in young people aged 13-18 years, those living in more affluent areas were more likely to be treated (13·8% in the lowest income category compared to 19·2% in the highest) (Schechter et al., 2011).

My second, and contrasting, set of findings on CF therapies, however, point to an apparent pro-rich bias in two other therapies, more evident in adults than children: more affluent adults in the UK are more likely to receive DNase and inhaled antibiotics than their more disadvantaged counterparts. DNase is a relatively expensive therapy to reduce viscosity of sputum and to aid sputum expectoration, and there is some evidence that it prevents decline in %FEV<sub>1</sub> (Pressler, 2008). These therapies, although relatively expensive, are free of charge to all patients in the NHS. One possibility for the social disparity in access to them is that they are both home-based treatments, requiring regular and long-term administration. Socially disadvantaged patients with CF are less likely to adhere to treatments (Schechter et al., 2009) and if they report poor adherence, clinicians might be less likely to prescribe these drugs because they are unlikely to be as effective if taken inconsistently. Evidence from the US shows no significant difference in use of DNase in children by area income quintile, but children on Medicaid (i.e. receiving free or subsidised care) were more likely to receive DNase than non-Medicaid children (Schechter et al., 2009).

Further research is needed to clarify which elements of the CF care model may contribute to a reduction in the adverse outcomes associated with deprivation. It is concerning that the burden of treatment for more disadvantaged families is higher and hospital admission to administer IV therapy, which is more disruptive to school and family life, is much more common. Furthermore the link with *P. aeruginosa* colonisation requires further investigation. Higher SES, as measured by parental education status, is associated with improved adherence to treatment in CF (Schechter, 2011), and further research to investigate the processes that lead to these

differences is required. Systems to support home IV therapy provision for more deprived groups in the UK should be explored.

Differences in access to health care cannot explain the differences in weight and height, by SES, that are evident at the time of diagnosis, and are unlikely to explain the gradient in lung function evident at around age five. The UK CF registry does not capture data about smoking in the home and these early effects may be associated with the known differences in smoking prevalence by SES in the UK (Mackenbach, 2011b). The effect of SES on growth *in utero* and in the early years in people with CF may be partially mediated by maternal smoking and ETS exposure, thus affecting subsequent outcomes and ultimately survival.

In conclusion, this study has identified important longitudinal inequalities in weight, height, BMI, lung function, and risk of *P. aeruginosa* carriage by deprivation in people with CF in the UK, which start early in life, but do not increase over time. The strength of the effect size, dose-response relationship, timing, consistency across outcomes, coherence with previous research, existence of plausible mechanisms, and absence of any obvious alternative explanation suggest that the associations are likely to be causal.

There is significant pro-poor bias in relation to IV therapies and nutritional supplementation, but evidence of reduced use of DNase and inhaled antibiotics after adjustment for need. Future studies should focus on interventions, such as reducing ETS exposure (Taylor-Robinson and Schechter, 2011), which may mitigate the effects of deprivation during the critical early years of life, and on identifying aspects of healthcare provision in CF that would help overcome the extra burden of adverse consequences of CF faced by patients living in disadvantaged circumstances.

### **Interpretation**

Referring to the Diderichsen model outlined in the literature review (Figure 6) this study has demonstrated *differential outcomes* in CF on the basis of SES. There are important longitudinal differences in weight, height, BMI, %FEV<sub>1</sub> and risk of *P. aeruginosa* colonisation (which could also be considered a *differential exposure*) by deprivation in people with CF in the UK, which start early in life, but do not increase over time. This analysis also suggests that there is *differential vulnerability*

(Figure 6) by SES, stemming from the same exposure to the underlying genetic risk factor for CF (e.g. two abnormal CFTR alleles). These pathways are discussed further in Chapter 7.

Marked socio-economic differences were observed in the reported use of key CF therapies in the UK. People from more deprived areas are around twice as likely to receive hospital IV antibiotic therapy, and nutritional support, but less likely to receive DNase and inhaled antibiotics. Interventions to reduce inequalities in outcomes in CF need to be focussed in the antenatal period and the early years. Further research is needed to clarify which elements of the CF care model in the UK may contribute to a reduction in the adverse outcomes associated with deprivation, and to investigate identified differences in access to inhaled therapies.

## **Chapter 5: Study 2 – A longitudinal study of the impact of social deprivation and disease severity on employment status in the UK CF population**

### **Abstract**

**Objective:** People with CF in the UK and elsewhere are increasingly surviving into adulthood, yet there is little research on the employment consequences of having CF. Poorer socio-economic circumstances have been linked with worse outcomes in CF. Exploring broader social patterns in employment in people with CF and other chronic illnesses is a key step in understanding how health and social inequalities are generated and perpetuated. I investigated, for the first time in a UK-wide cohort, longitudinal employment status, and its association with deprivation, disease severity, and time in hospital.

**Design:** I undertook a longitudinal registry study of adults with CF in the UK aged 20 to 40 (3458 people with 15,572 observations between 1996 and 2010).

**Methods:** Mixed-effects models were used to assess the association between small area deprivation and employment status, adjusting for clinically important covariates.

**Results:** Around 50% of adults with CF were in employment. Male sex, higher %FEV<sub>1</sub> and BMI, and less time in hospital were associated with improved employment chances. A person in the most deprived quintile was less likely to be in employment, after adjusting for disease severity, compared to their more advantaged counterparts (log-odds -2.66, 95% CI -3.1 to -2.26, comparing the most to the least deprived quintile). For men with a %FEV<sub>1</sub> of 60 at the age of 30, this equates to 67.7% employment in the least deprived quintile, compared to 50.2% in the most deprived, a difference of 17.5 percentage points. Comparing a population with relatively good lung function (a %FEV of 80), to one with poor lung function (a %FEV<sub>1</sub> of 30), with all other things being equal (i.e. deprivation quintile 3, male sex), at the age of 30 there was a difference of 7.2 percentage points in employment chances. Furthermore, deprivation modifies the effect of lung function on employment chances – a 40 point contrast in %FEV<sub>1</sub> was associated with a three percentage point reduction in employment chances in the most advantaged quintile,

compared to a 7.7 point reduction in the most deprived quintile. Genotype and use of home IV therapy were not associated with employment status.

**Conclusions:** Deprivation, disease severity, and time in hospital all influence employment chances in CF. Furthermore, our analysis demonstrates that deprivation amplifies the harmful effects of disease severity on employment: the employment chances of people with CF with poor lung function from disadvantaged areas are damaged to a much greater extent than for their counterparts living in the least disadvantaged circumstances. This *differential social consequence* of having CF is likely to be a key pathway for the amplification of health inequalities in CF. Future studies should focus on understanding and mitigating the barriers to employment faced by people with CF.

## **Introduction**

CF is the commonest life-limiting inherited disease among Caucasian populations, with most patients dying prematurely from respiratory failure. Children with CF in the UK and other OECD countries are usually diagnosed in the first year of life (Cystic Fibrosis Trust, 2011), and subsequently require intensive support from family and healthcare services. People with CF in the UK and elsewhere are increasingly surviving into adulthood, with the median age of survival estimated to be over 50 years for a person born in this century (Dodge et al., 2007). An implication of this is that increasing attention needs to be paid to the experiences of people with CF in adulthood, including the employment consequences of having CF.

Employment is one of the ‘social determinants’ of health (Marmot et al., 2010b). Work influences health in a number of ways; it provides income to meet material needs, but also fulfils critical psycho-social functions, increasing self-worth, sense of identity and opportunities for social interaction. Numerous studies have identified unemployment as a potent risk factor for poor health, and equally, poor health can lead to reduced employment chances (Bambra et al., 2009, Bambra et al., 2010). People with chronic illnesses face numerous barriers to entering the labour market, and CF provides a case in point. Factors related to disease severity, such as reduced lung function may restrict employment choices for adults with CF, and the treatment burden further compounds this; adults with CF are generally expected to perform physiotherapy regularly and there are the added demands of taking large numbers of therapies, including frequent visits to hospital (Sawicki et al., 2009).

Despite this, the evidence about patterns of employment for adults with CF is limited (Saldana and Pomeranz, 2012), and mainly based on cross-sectional studies of single centres, which cannot delineate the relationships between chronic illness and employment outcomes or whether these relationships indicate causality. Furthermore, Edwards et al (Edwards and Boxall, 2010), adopting the social model of disability, have criticised the approach taken to understanding employment outcomes in CF. They point out that most of the research, to date, focusses on the effects of disease severity on employment chances. They argue that this ‘impairment’ focus locates the problem within the individual, and ignores the significant structural and societal barriers to employment for people with chronic illness.

Exploring broader social patterns in employment in people with CF and other chronic illnesses is a key step in understanding how health and social inequalities are generated and perpetuated. CF provides a particularly useful case for studying these processes, because it is a classically inherited genetic disease, and unlike most chronic diseases, there is no difference in the incidence of the condition with SES (Taylor-Robinson and Schechter, 2011, Taylor-Robinson et al., 2013a). However, inequalities develop over the course of people's lives, as a consequence of having the disease. Informed by Diderichsen's analytic framework of the pathways from social context to health outcomes (Diderichsen et al., 2001), I have demonstrated important differential health consequences of having CF, by deprivation in the UK (see Study 1, Chapter 4). For instance, in the UK there are clinically important differences in growth, and lung function by deprivation, which are evident early on in children's lives (Taylor-Robinson and Schechter, 2011, Taylor-Robinson et al., 2013a). Furthermore, people with CF from socio-economically disadvantaged backgrounds die at a younger age than those in more advantaged social positions in the UK and the US (Britton, 1989, Schechter et al., 2001, O'Connor et al., 2003, Barr HL, 2011). The social patterning of outcomes in CF implies that *differential exposure* to social and environmental risk factors is playing an important role in influencing outcomes (Diderichsen et al., 2001, Ben-Shlomo and Kuh, 2002, Marmot et al., 2010b).

Building on these findings, the next step is to look for any 'differential social consequences' of ill-health in the context of CF (Diderichsen et al., 2001). My aims in this study were to explore the effect of deprivation, disease severity, and time in hospital on longitudinal employment chances in people with CF and to investigate whether changes in lung function have differential effects on employment chances by deprivation ('*differential social consequences*' in *Diderichsen's model*, Figure 6 in Chapter 3). For instance, is low lung function in CF more damaging to employment chances in people from more disadvantaged areas? I undertook a longitudinal population level registry study of employment status in adults with CF in the UK to answer such questions.

## **Methods**

An in depth description of the methods used is found in Chapter 3. An overview is provided here.

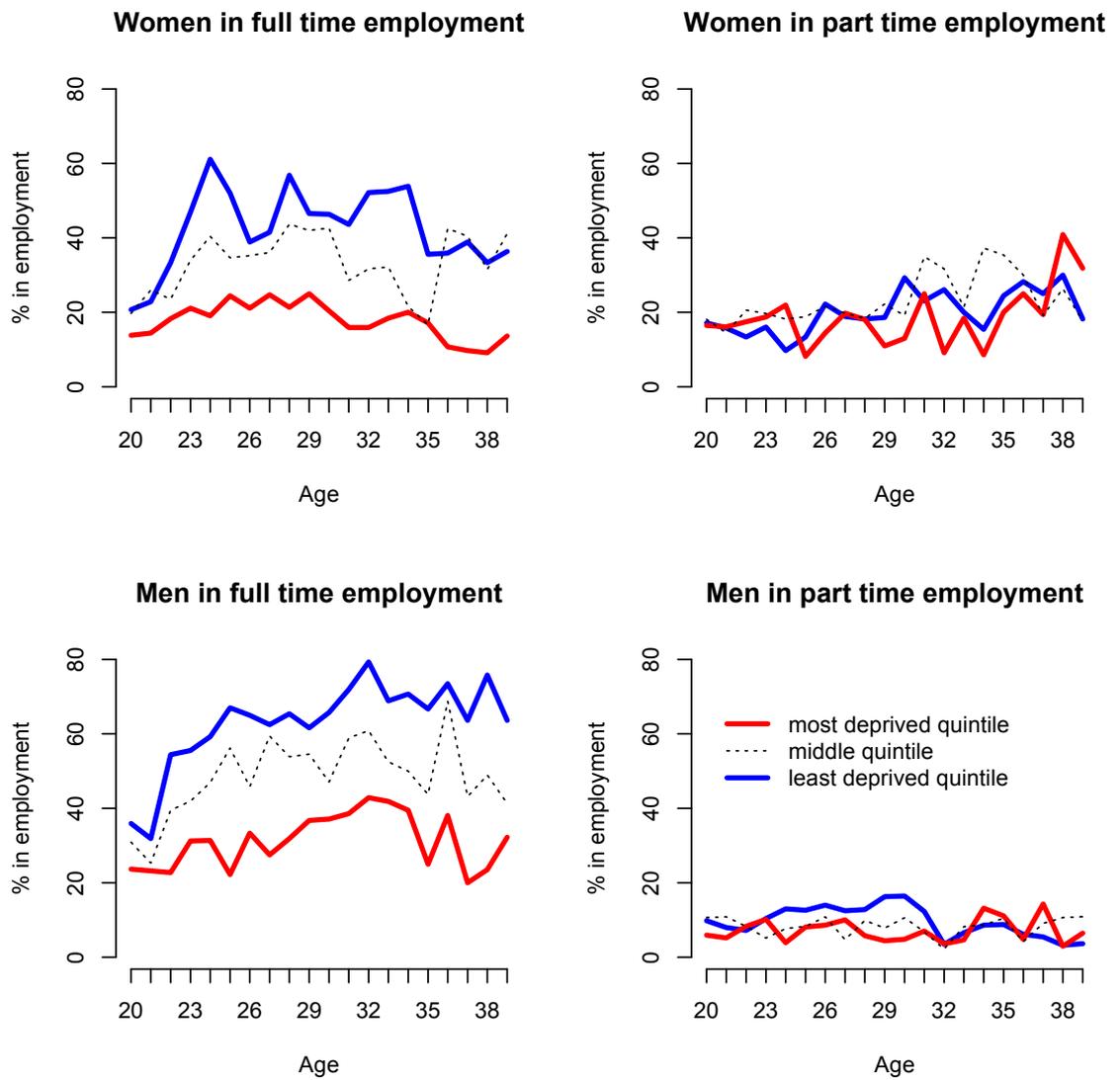
### ***Design, setting and data source***

I undertook a longitudinal retrospective cohort study of annual review data on individuals between the ages of 20 and 40 with at least one outcome measurement and a valid postal code in the UK CF Registry between 1996 and 2010. The UK CF Registry is supported and coordinated by the UK CF Trust (CF Trust, 2013a). The Registry is maintained to a high standard of data quality, and is estimated to include nearly all people estimated to have CF in the UK population (Mehta et al., 2004) and is therefore ideally suited to the study of outcomes across the whole socio-economic spectrum in the UK society.

### ***Primary outcome and covariates***

The primary outcome was any employment in the preceding year (yes or no), which included people recorded as being in either full- or part-time employment at annual review (Figure 57). The primary exposure measure was a small-area-based measure of deprivation of area of residence. Postcodes were used to derive IMD scores for the constituent UK countries (GeoConvert, 2011) and each person was allocated to a deprivation quintile on the basis of first recorded postcode. Baseline covariates in the analysis were: sex; genotype coded as the number of delta F508 alleles (0, 1 or 2); and year of birth. I adjusted for disease severity in the analysis on the basis of lung function (measured by %FEV<sub>1</sub>) and body mass (measured by BMI SD score). %FEV<sub>1</sub> is recognized as key outcome measure in CF, as it is strongly predictive of survival (Rosenfeld et al., 2005). As a measure of time spent administering therapies, we included the number of IV therapy days in the past year in our analysis, further disaggregated into IV days in hospital, and IV days at home.

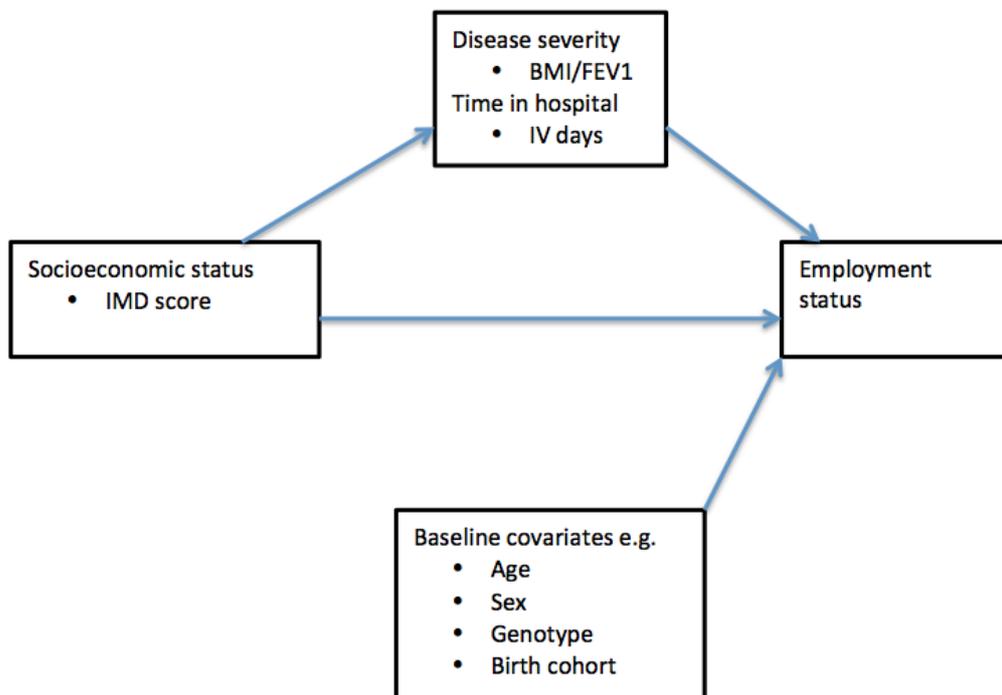
**Figure 57: Cross-sectional full time and part time employment, by deprivation and sex**



### *Statistical Methods*

In brief, exploratory analysis involved: plotting the stratified raw data; fitting GAMs (Hastie and Tibshirani, 1990) to visualize the shape of associations; and plotting empirical logits (log-odds of employment binned by age) from the raw data. We then fitted GLMMs to the data across the age range. These model the log-odds of employment status as a linear function of the measured covariates and individual level random-effects (see methods chapter). Linear and quadratic models for the mean trajectory were explored as appropriate, as informed by the GAMs. The logic model for the analysis is shown in Figure 58. I first fitted a model adjusted for age and the baseline covariates defined above, which are unlikely to be in the causal path from SES to employment status. I then tested for the significance of adding disease severity measures, and service use measures, and finally added deprivation score to the model.

**Figure 58 Logic model to inform analysis of employment status**



In this way I fitted sequential models adjusting for the covariates of interest, and estimated model parameters by maximum likelihood, using generalized likelihood ratio statistics to compare nested models, and Wald statistics to test hypotheses about model parameters (Laird and Ware, 1982). These longitudinal models take into account drop-out due to death, and implicitly estimate the chances of employment in a drop-out free population (Diggle et al., 2002). I present effect estimates as log-odds with CIs, since ORs can be misinterpreted when outcomes are common (Grimes and Schulz, 2008). To aid interpretation, I display population-averaged employment chances in the plots, by averaging individual-level fitted values over the population.

### *Ethics*

NHS research ethics approval (Huntingdon Research Ethics Committee 07/Q0104/2) has been granted for the collection of data into the UK database. The CF Trust database committee approved the use of anonymised data in this study.

## Results

### Population characteristics

In total, 4062 people were recorded in the registry between the ages of > 20 and < 40 years; 3495 of these had a valid postcode recorded and met the inclusion criteria and 3458 individuals had full data on all baseline covariates. The final analysis dataset thus contained 3458 people, with 15,098 person-years of follow-up, and data collected at 15,572 annual reviews. The baseline characteristics of individuals at first recorded entry into the cohort, are shown in Table 13, stratified by employment status, and deprivation quintile (Table 14). At first entry to the cohort, 53% of people were in full-time employment. Of those in employment 14% were in the most deprived quintile, while of those not in employment 25% were in the most deprived quintile.

**Table 13: Characteristics of study population in UK CF Registry by employment status at baseline**

	Not in employment	Employed	Total
Number of adults with CF (%)	1845 (53.4)	1613 (46.6)	3458
Observations (%)	7287 (46.8)	8285 (53.2)	15572
Deprivation quintile 1 (least deprived)	295 (16)	344 (21.3)	639 (18.5)
Deprivation quintile 2	319 (17.3)	370 (22.9)	689 (19.9)
Deprivation quintile 3	357 (19.3)	354 (21.9)	711 (20.6)
Deprivation quintile 4	417 (22.6)	321 (19.9)	738 (21.3)
Deprivation quintile 5 (most deprived)	457 (24.8)	224 (13.9)	681 (19.7)
Number of F508 alleles:2 (%)	952 (51.6)	744 (46.1)	1696 (49)
Number of F508 alleles:1 (%)	632 (34.3)	616 (38.2)	1248 (36.1)
Number of F508 alleles:0 (%)	261 (14.1)	253 (15.7)	514 (14.9)
Female	856 (46.4)	672 (41.7)	1528 (44.2)
Non-white	54 (2.9)	27 (1.7)	81 (2.3)
Birth cohort >1959-01-01	152 (8.2)	211 (13.1)	363 (10.5)
> 1969-01-01	409 (22.2)	573 (35.5)	982 (28.4)
>1979-01-01	1203 (65.2)	780 (48.4)	1983 (57.3)
>1989-01-01	81 (4.4)	49 (3)	130 (3.8)
Median age at baseline (years) (IQR)	21 (20.4,24.5)	23 (20.7,29.3)	21.5 (20.5,27)
Median %FEV <sub>1</sub> at entry (IQR)	61.8 (41.8,82.2)	68.4 (49.9,85.2)	65.3 (45.8,83.7)
%FEV <sub>1</sub> >90 (normal)	286 (15.5)	302 (18.7)	588 (17)
%FEV <sub>1</sub> >70 and <90 (mild)	457 (24.8)	458 (28.4)	915 (26.5)
% FEV <sub>1</sub> >40 and <70 (moderate)	678 (36.7)	619 (38.4)	1297 (37.5)
% FEV <sub>1</sub> <40 (severe)	424 (23)	234 (14.5)	658 (19)
Pseudomonas colonization at entry	1123 (60.9)	836 (51.8)	1959 (56.7)
Median BMI SDS at entry (IQR)	-0.6 (-1.4,0.1)	-0.4 (-1.2,0.4)	-0.5 (-1.3,0.3)
Died	223 (12.1)	115 (7.1)	338 (9.8)

**Table 14: Baseline characteristics at entry to the dataset, stratified by deprivation quintile**

	<i>1 (least deprived)</i>		<i>2</i>		<i>3</i>		<i>4</i>		<i>5 (most deprived)</i>		<i>All</i>
Number of people (%)	641 (18.5)	689 (19.9)	711 (20.5)	738 (21.3)	682 (19.7)	3461					
Observations (%)	2940 (18.2)	3346 (20.7)	3427 (21.2)	3448 (21.3)	3001 (18.6)	16162					
Employment at baseline (%)	345 (53.8)	370 (53.7)	353 (49.6)	320 (43.4)	224 (32.8)	1612 (46.6)					
Age at diagnosis (days)	249 (33,1461)	183 (33,998.2)	253.5 (49.5,1154.8)	199 (37.5,1233.5)	193 (61,1526)	213 (48,1279)					
Number of F508 alleles <sub>2</sub> (%)	332 (51.8)	359 (52.1)	348 (48.9)	363 (49.2)	296 (43.4)	1698 (49.1)					
Number of F508 alleles <sub>1</sub> (%)	231 (36)	243 (35.3)	246 (34.6)	264 (35.8)	264 (38.7)	1248 (36.1)					
Number of F508 alleles <sub>0</sub> (%)	78 (12.2)	87 (12.6)	117 (16.5)	111 (15)	122 (17.9)	515 (14.9)					
Female (%)	268 (41.8)	309 (44.8)	301 (42.3)	337 (45.7)	315 (46.2)	1530 (44.2)					
Non-white (%) <sub>...</sub>	7 (1.1)	6 (0.9)	12 (1.7)	21 (2.8)	35 (5.1)	81 (2.3)					
Screened (%)	46 (7.2)	52 (7.5)	34 (4.8)	44 (6)	30 (4.4)	206 (6)					
Birth cohort >1959-01-01	89 (13.9)	69 (10)	80 (11.3)	67 (9.1)	58 (8.5)	363 (10.5)					
>1969-01-01	175 (27.3)	199 (28.9)	219 (30.8)	211 (28.6)	178 (26.1)	982 (28.4)					
>1979-01-01	353 (55.1)	394 (57.2)	382 (53.7)	430 (58.3)	426 (62.5)	1985 (57.4)					
>1989-01-01	24 (3.7)	27 (3.9)	30 (4.2)	30 (4.1)	20 (2.9)	131 (3.8)					
Median age at baseline (years) (IQR)	21.6 (20.5,28.7)	21.3 (20.5,27)	22 (20.6,27.3)	21.6 (20.5,26.6)	21.2 (20.6,25.9)	21.5 (20.5,27)					
Median %FEV1 at entry (IQR)	65.7 (46.4,83.7)	67 (44.7,84)	65.4 (46.9,83.4)	64.2 (45.5,83.5)	64.5 (45.5,83.8)	65.3 (45.8,83.7)					
Pseudomonas colonization at entry	341 (53.2)	412 (59.8)	393 (55.3)	399 (54.1)	413 (60.6)	1958 (56.6)					
Median BMI SDS at entry (IQR)	-0.4 (-1.1,0.4)	-0.4 (-1.2,0.2)	-0.5 (-1.3,0.3)	-0.6 (-1.4,0.2)	-0.6 (-1.4,0.2)	-0.5 (-1.3,0.3)					

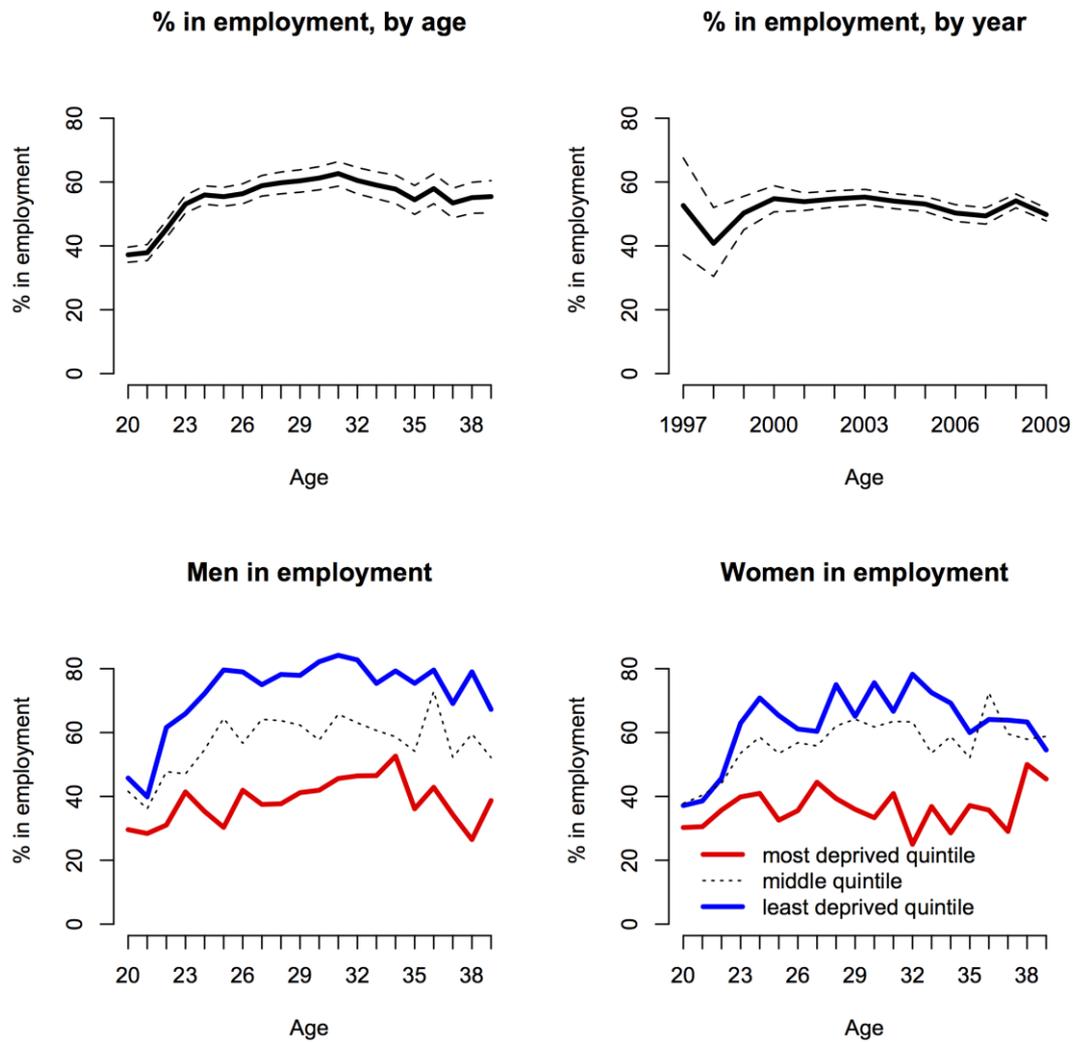
## **Exploratory analysis**

In each annual data collection wave in the registry, for all ages between 20 and 40, about 50% of the UK CF population were recorded as being in full- or part-time employment (Figure 59). When stratified by age, employment prevalence increased from 37% at age 20, before peaking at around 60% at age 30, and subsequently decreasing to 55% at age 40 (Figure 59). These overall trends hide important differences in employment prevalence by sex and deprivation status. As illustrated in the lower row of Figure 59, the cross-sectional prevalence of employment was greater in people from the least deprived quintile at all ages. In the least deprived areas, the prevalence of full-time employment increased to around 80% in men and 70% in women, between the ages of 20 and 25 years, before decreasing gradually after the age of 30 years. By contrast, in people from the most deprived quintile, the prevalence of employment remained fairly constant, at around 30% to 40% for both men and women.

The GAMs (Figure 60) suggested an approximately linear relationship between deprivation score and log-odds of employment, with increasing deprivation associated with decreased employment chances. Increasing %FEV<sub>1</sub> and increasing age were associated with increased employment chances, and the relationship with birth cohort did not change over time. On this basis, birth cohort was included in the model as a continuous variable, rather than a factor. The empirical logit plots suggested important differences in employment chances by SES, and that a quadratic term may be necessary to model the age trajectory (Figure 61).

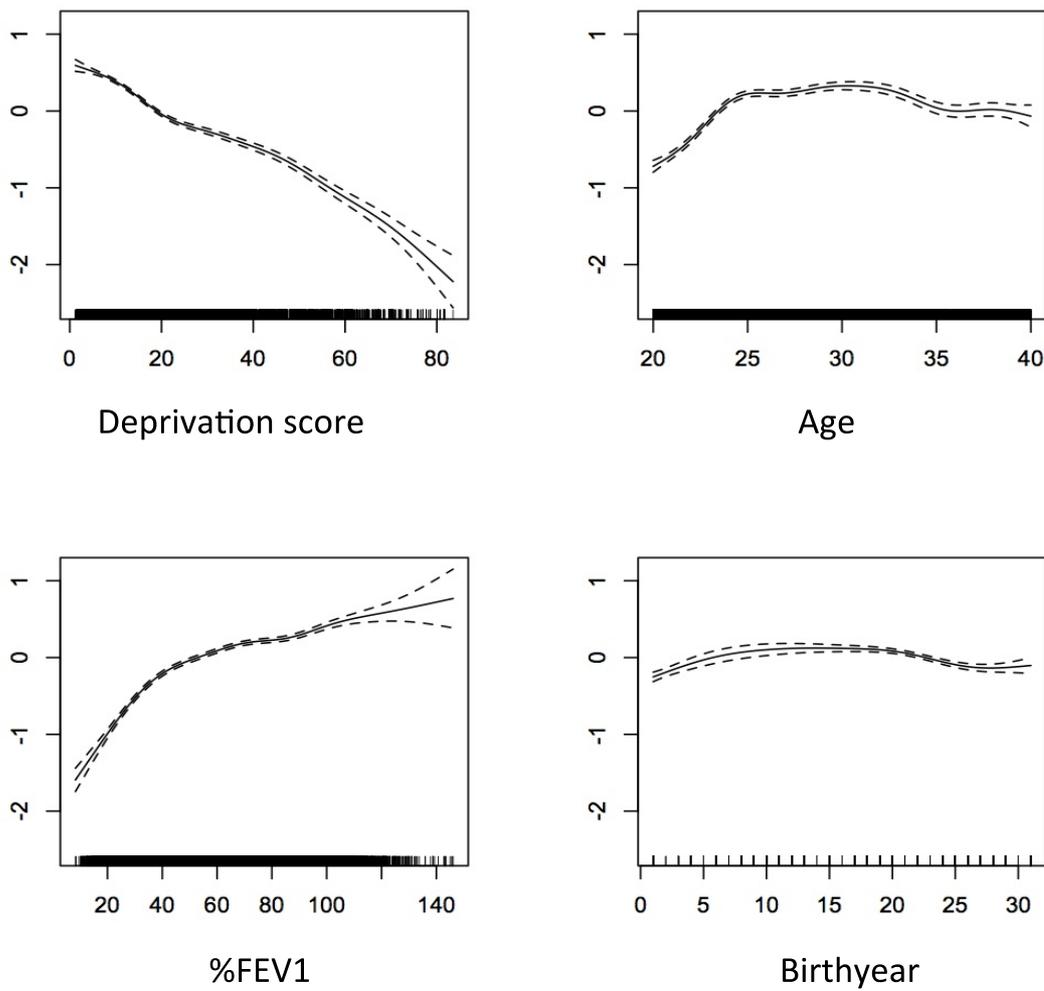
**Figure 59: Cross sectional employment prevalence by age and year**

With 95% CIs (top row). Bottom row shows cross-sectional employment prevalence by age, stratified by deprivation quintile (most deprived quintile in red), for men and women.

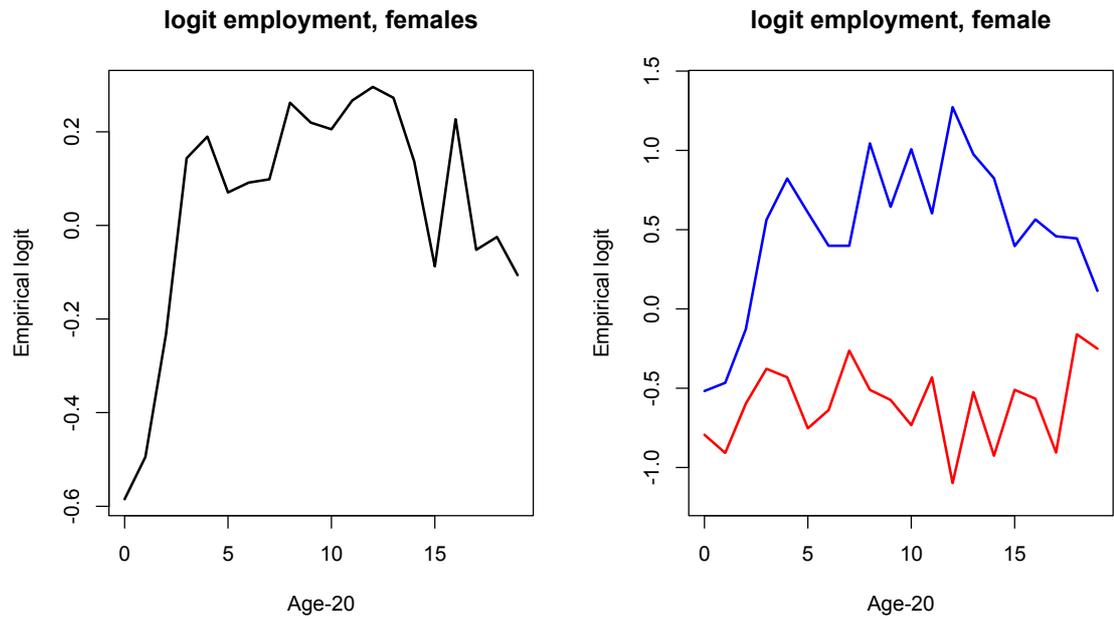


**Figure 60: GAMs for risk of employment**

GAMs showing the mutually adjusted cross-sectional effect of deprivation score (top left), age (top right), %FEV<sub>1</sub> (bottom left) and (birthyear-1959) (bottom right) on scaled risk of employment. Y axis in the plots shows log-odds of employment.



**Figure 61: Empirical logit plots by age and SES for females in the dataset**



## Definitive analysis

Figure 62 illustrates the modelled independent population averaged effects of deprivation, sex, %FEV<sub>1</sub>, BMI and time in hospital on employment chances for people with CF in the UK, on the basis of the final model (Table 15, column 5). There are significant age-related effects. The general pattern is for the proportion of people in employment to increase to around age 30, and decrease subsequently. Genotype and use of home IV therapy were not associated with employment status in any of the analyses, and I excluded them from the final models.

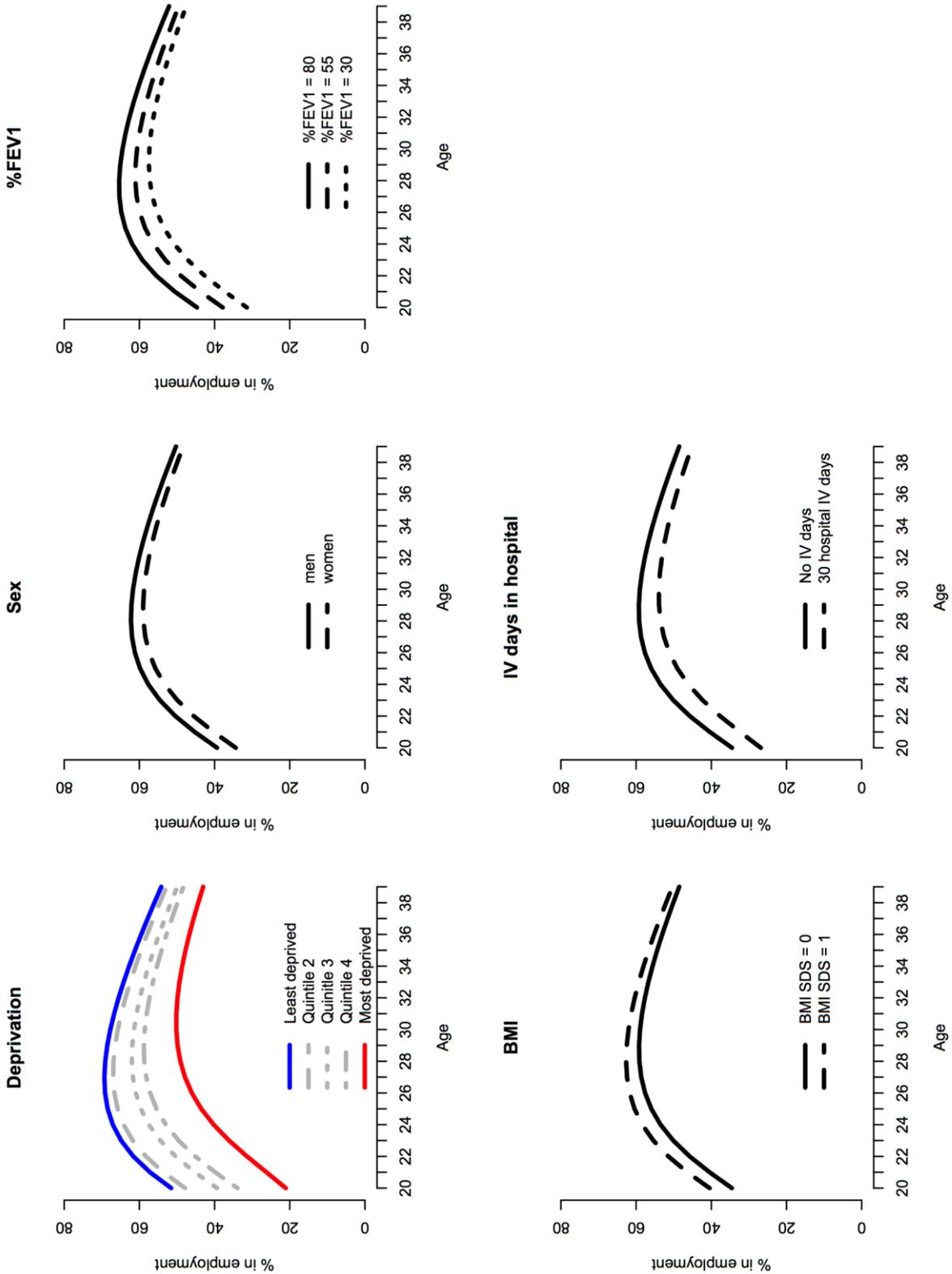
Of the covariates in the model, deprivation status explained more of the variance, and there was a dose-response relationship (Figure 62, Table 15): a person in the most deprived quintile was less likely to be in employment, after adjusting for disease severity, compared to their more advantaged counterparts (log-odds -2.66, 95% CI -3.1 to -2.26, comparing the most to the least deprived quintile). For men with a %FEV<sub>1</sub> of 60 at the age of 30, this equates to 67.7% employment in the least deprived quintile, compared to 50.2% in the most deprived, a difference of 17.5 percentage points (Figure 62). Men were more likely than women to be in employment (log-odds 0.40, 95% CI 0.16 to 0.64 in adjusted model), which corresponds to 61.7% employment in men, compared to 58.7% employment in women at age 30, for people with a %FEV<sub>1</sub> of 60, in the middle deprivation quintile – a difference of three percentage points. People with better lung function were more likely to be in employment, and this followed a monotone dose response relationship. For a population with relatively good lung function (a %FEV<sub>1</sub> of 80) at age 30, this equates to 64.5% employment, compared to 57.3% in a population with relatively poor lung function (a %FEV<sub>1</sub> of 30), with all other things being equal (i.e. deprivation quintile 3, male sex) (Figure 62). Higher BMI was associated with improved employment chances (log-odds 0.1, 95% CI 0.020 to 0.188 per one unit increase in BMI SD score), and more days in hospital were associated with lower employment chances (log-odds -0.023, 95% CI -0.027 to -0.019, per day in hospital).

**Table 15: Log-odds for the final nested GLMMs.**

	<i>Baseline</i>	<i>Baseline + severity</i>	<i>Baseline + Severity + Time in hospital</i>	<i>Baseline + Severity + Time in hospital + Deprivation</i>	<i>Baseline + Severity + Time in hospital + Deprivation*%FEV<sub>1</sub></i>
Constant	0.663*** (0.105)	0.761*** (0.106)	0.997*** (0.105)	2.047*** (0.172)	2.058*** (0.171)
age	0.128*** (0.016)	0.143*** (0.016)	0.144*** (0.016)	0.146*** (0.016)	0.146*** (0.016)
age^2	-0.022*** (0.002)	-0.024*** (0.002)	-0.023*** (0.002)	-0.023*** (0.002)	-0.023*** (0.002)
Birthyear	-0.057*** (0.014)	-0.059*** (0.014)	-0.046** (0.014)	-0.039** (0.014)	-0.039** (0.014)
Male/Female	0.446*** (0.129)	0.485*** (0.128)	0.443*** (0.125)	0.410** (0.125)	0.401** (0.125)
Random intercept SD	(3.016)	(2.893)	(2.779)	(2.640)	(2.632)
Random slope SD	(0.450)	(0.451)	(0.438)	(0.442)	(0.441)
%FEV <sub>1</sub>		0.026*** (0.002)	0.020*** (0.002)	0.021*** (0.002)	0.013** (0.005)
BMI SDS score		0.163*** (0.043)	0.121** (0.043)	0.106* (0.043)	0.104* (0.043)
Hospital IV days			-0.024*** (0.002)	-0.023*** (0.002)	-0.023*** (0.002)
Deprivation quintile 2/1				-0.270 (0.202)	-0.279 (0.202)
Deprivation quintile 3/1				-0.967*** (0.200)	-0.976*** (0.200)
Deprivation quintile 4/1				-1.422*** (0.198)	-1.427*** (0.198)
Deprivation quintile 5/1				-2.650*** (0.207)	-2.663*** (0.207)
Deprivation quintile 2/1 x %FEV <sub>1</sub>					0.010 (0.007)
Deprivation quintile 3/1 x %FEV <sub>1</sub>					0.010 (0.006)
Deprivation quintile 4/1 x %FEV <sub>1</sub>					0.006 (0.006)
Deprivation quintile 5/1 x %FEV <sub>1</sub>					0.016* (0.007)
Log-likelihood	-7885	-7720	-7645	-7548	-7545
N	15572	15430	15430	15430	15430
Groups	3458	3451	3451	3451	3451
Baseline variance explained (%)	-	7.9	15.1	23.3	23.8

\*P < 0.05, \*\* P < 0.01, \*\*\* P < 0.001  
Standard errors in parentheses

**Figure 62. Longitudinal employment trajectory versus age of people with CF in UK CF Registry, by deprivation quintile, sex, %FEV<sub>1</sub>, BMI SD score and days in hospital.**



In order to assess if there was an interaction between deprivation status and %FEV<sub>1</sub> level, I first undertook an analysis of individuals in deprivation quintiles 1 and 5 only, and dichotomised these patients into those with high and low lung function, using a %FEV<sub>1</sub> of 50 percentage points as the cut off. Figure 63 shows the raw data plotted by age, and this provided some evidence of an interaction, with an apparent larger effect of low lung function on employment chances evident in people in the most deprived quintile. Formally fitting a model to this data, with covariates limited to age, deprivation quintile, and dichotomised %FEV<sub>1</sub> also showed a significant interaction (Figure 63).

**Figure 63: Exploratory analysis of interaction effect between deprivation and %FEV<sub>1</sub> on employment status.**

*Raw data (left panel) and fitted model (right panel) limited to deprivation quintile 1 and 5, and %FEV<sub>1</sub> dichotomised into high (>50) and low categories (<50).*

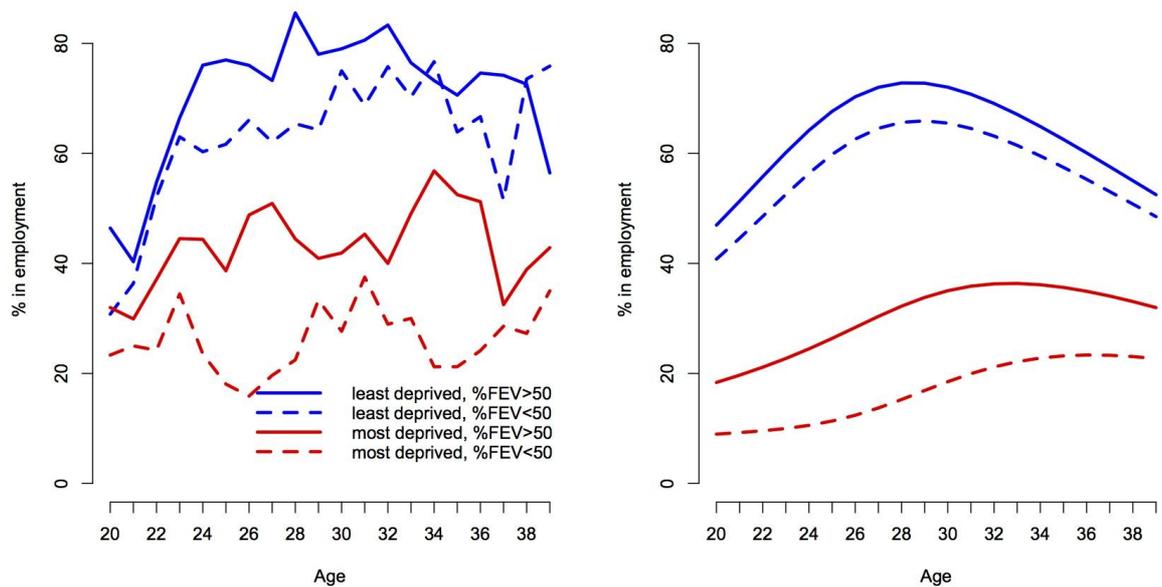


Figure 64 illustrates the interactive effect of level of lung function and social deprivation on population averaged employment chances in men and women, in the final model fitted to all of the data. In the final model, a composite test for interaction between %FEV<sub>1</sub> and deprivation quintile is not significant (P = 0.2), but the contrast between the most deprived quintile and %FEV<sub>1</sub> is significant (log-odds 0.016, 95% CI 0.0028 to 0.03, per unit increase in %FEV<sub>1</sub>, P = 0.018) (Table 15). The fitted data for men and women at age 30 is shown in Table 16. This indicates that declines in lung function have more of an effect on employment chances in disadvantaged populations – at the age of 30, a contrast of 40 percentage points in lung function corresponds to a decline in employment prevalence from 69.2% to 66.1% (3.1 percentage points difference) for men in least deprived quintile, compared to a change from 54.1% to 46.4% (7.7 percentage points difference) in the most deprived quintile (Table 16).

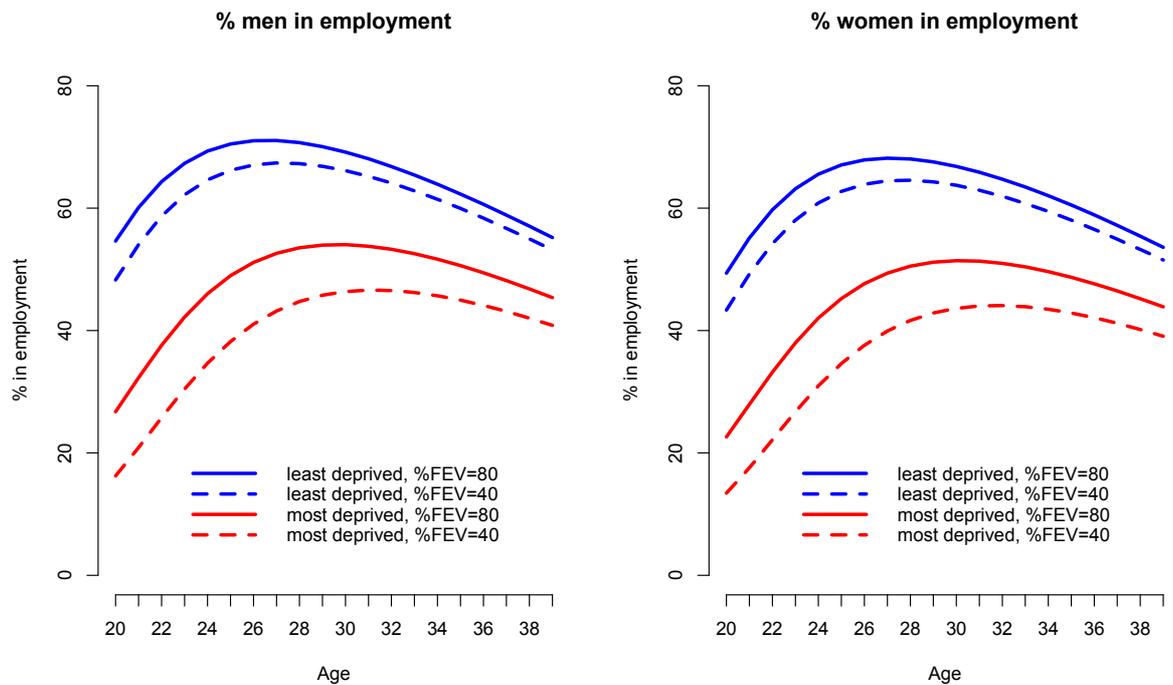
**Table 16: Percentage of people with CF in employment at age 30**

	Employment (%) MEN	Employment (%) WOMEN
No health or social disadvantage (good lung function and high SES)	69.2	66.8
Health disadvantage only (poor lung function and high SES)	66.1	63.7
Social disadvantage only (good lung function and low SES)	54.1	51.4
Double disadvantage (poor lung function and low SES)	46.4	43.6

*Data from final interaction model  
 %FEV<sub>1</sub> fixed at 80 percentage points for good lung function, and 40 percentage points for poor lung function  
 High SES fixed at least deprived quintile, and low SES fixed at most deprived quintile*

**Figure 64: Longitudinal employment trajectory versus age, demonstrating the interaction between deprivation and %FEV<sub>1</sub>**

*The lines show the final modelled longitudinal trajectories from the interaction model (Table 15, column 5), contrasting the adjusted effects of deprivation, and %FEV<sub>1</sub>. Low %FEV<sub>1</sub> is more damaging to employment chances in people living in the most disadvantaged areas. Other covariates in the final model are fixed.*



## Discussion

I undertook a longitudinal registry-based study of employment status in the UK CF population, and found that deprivation status is a more important predictor of employment chances than either disease severity or time spent in hospital. Furthermore, deprivation modifies the effect of lung function on employment chances. Our analysis demonstrates that there are *differential social consequences* of poor health in the context of CF – poor lung function is approximately twice as harmful to employment chances in people living in the most deprived areas, compared to the least. As people are living longer, healthier lives with CF, it is more important than ever for strategies to promote employment to focus on the broader societal barriers to engagement in the workforce for people with CF, rather than taking a narrow ‘impairment’ focus solely on the impact of disease severity on employment chances, as critiqued by Edwards et al (Edwards and Boxall, 2010).

Key strengths of this study include the population-wide coverage of the UK CF registry, the high quality of the data, the longitudinal analysis, and the theoretical approach that responds to previous criticisms of the illness-focussed approach to understanding employment outcomes in CF. These findings are particularly relevant to the UK population, but could be cautiously generalized to other OECD countries with an universal health service. There are limitations: first it relies on retrospective, routinely collected data and we used a standard measure of deprivation of area of residence. It was therefore not possible to separate effects of socio-economic circumstances operating at the individual and area level. Second, I only had valid postcodes on 86% of the sample, though our sample size is large, with no pronounced gradient in the proportion of patients with valid postcodes by deprivation quintile i.e. missing postcode information is unrelated to deprivation.

A recent review concluded that further research on CF and employment is necessary to improve occupational outcomes (Saldana and Pomeranz, 2012). The systematic review identified nine studies, that have looked at the relationship between CF and employment status, all of which were small, and based at one or two care centres. Six studies reported employment rates in all of these the employment rate was around 50%. This review did not include the largest study to date, by Walters et al (Walters et al., 1993), which was a cross-sectional questionnaire survey of 1052 adults over 16

years of age with CF in the UK in 1990. Walters et al found that 55% of responders were working, whilst of those not employed, half gave ill health as the reason (Walters et al., 1993). In our study we find that the annual employment rate among CF adults in the UK appears to have remained unchanged between 50% and 55% over the last decade. Five previous studies focussed on the association between disease severity, as measured by %FEV<sub>1</sub>, and employment status, with mixed results: three studies conducted in the US (Burker et al., 2004), Canada (Frangolias et al., 2003) and Australia (Hogg et al., 2007) concluded that %FEV<sub>1</sub> was not related to employment status, and two from the US (Gillen et al., 1995) and Belgium (Havermans et al., 2009) suggested that it was related. The studies by Burker et al. (Burker et al., 2004), Gillen et al. (Gillen et al., 1995), and Hogg et al. (Hogg et al., 2007) suggested that frequency of hospital admissions, demographic variables, mental health, and education level, may also influence employment chances.

This study suggests that disease severity and time in hospital also influence employment chances in the UK, but these effects are not as large as one might have predicted. It is evident that people with significant respiratory impairment continue to work, and disease severity alone does not predict employment outcomes. Furthermore, there is a large amount of variability between individuals with similar characteristics (large random effects), which suggests that there are other important factors related to employment status that we have been unable to account for in our study. Highest educational attainment is one such factor, but these data were available only for 60% of the individuals in our analysis (see Appendix 4, Table 41). Adjusting for educational attainment did not change any of the substantive effects in our analysis, though it did reduce the random effects variance further. Other studies have also suggested that individual psychological factors and education status are correlated with employment status in CF (Burker et al., 2004).

These previous studies on employment chances in people with CF tend to portray CF as a 'serious illness', which causes employment problems. In contrast, Edward et al explored the employment experiences of adults with CF from a social model of disability perspective. They demonstrated barriers to employment that were similar to those experienced by other disabled people, as well as barriers related to the 'impairment effects' of CF, and concluded that adults with CF have valuable perspectives to contribute to social model analysis and the development of

employment-related policy solutions (Edwards and Boxall, 2010). Our results not only corroborate, but also extend these observations, by demonstrating the interaction between disease severity related factors, and deprivation.

These findings add to the extensive literature on the inter-relationship between chronic illness, SES, and employment opportunities (Holland et al., 2011b). In the UK in 2005 the age standardized employment rate for people of working age (25-59) was 80% in healthy women, compared to 50% in those with limiting long-standing illness (LLSI), and 93% compared to 59% in men (Holland et al., 2011b). Furthermore there was a striking social gradient for those with LLSI – the prevalence of employment was 66% in highly educated women with LLSI compared to 18% in those with low education, and in men 72% compared to 30% (Holland et al., 2011a). In our study the cross-sectional prevalence of employment was around 60% in the most affluent quintile, compared to 30% in the most deprived quintile in women, and 70% versus 30% in men (Figure 57).

Low employment in people with CF is a serious concern. Despite there being no difference in incidence of CF by SES, there are important differences in outcomes such as growth and lung function, and ultimately survival, in people with CF by SES in the UK and US (Schechter et al., 2001, Barr HL, 2011, Taylor-Robinson et al., 2013a). Being out of work increases the risk of poverty and social exclusion, and is likely to further damage the health of the most disadvantaged people with CF. In this study I have demonstrated *differential social consequences of illness* in the context of CF, by which people with the double burden of chronic illness and low SES are more likely to be excluded from the labour market. We speculate that this may be an important pathway for the amplification of health inequalities in CF, whereby disadvantage builds on disadvantage. It is of particular concern that the most disadvantaged women have the poorest employment chances, since female sex is also an important risk factor for poor survival in CF (Barr HL, 2011).

In conclusion, this study has identified important longitudinal inequalities in employment outcomes by deprivation in people with CF in the UK, which particularly effect women living in the most disadvantaged areas. Future studies should focus on policy interventions that would help overcome the extra burden of adverse consequences of CF faced by patients living in disadvantaged circumstances.

## **Interpretation**

People with CF are living longer so it is important to understand factors that influence adult employment chances. Previous studies have shown that people with CF from socio-economically disadvantaged backgrounds have worse outcomes. This longitudinal study explores how SES, disease severity, and time in hospital influence employment chances, in order to better understand how health and social inequalities in CF are generated and perpetuated.

Around 50% of adults with CF were in employment. Lower social deprivation, male sex, higher %FEV<sub>1</sub> and BMI, and less time in hospital were associated with improved employment chances. Crucially, our analysis demonstrates for the first time that deprivation modifies the effect of disease severity in CF: poor lung function is more harmful to employment chances for people living in the most disadvantaged circumstances compared to the least. *Differential social consequences* (see Diderichsen model in Chapter 2, Figure 6) of illness, demonstrated here in the context of CF, are important in the generation of health inequalities. Policies should address the extra burden of adverse consequences of CF faced by patients living in disadvantaged circumstances.

## **Chapter 6: Study 3 – Understanding the natural progression in %FEV<sub>1</sub> decline in patients with CF: A longitudinal study**

### **Abstract**

#### **Background**

Forced expiratory volume in one second as a percentage of predicted (%FEV<sub>1</sub>) is a key outcome in CF and other lung diseases. As people with CF survive for longer periods, new methods are required to understand the way %FEV<sub>1</sub> changes over time. We present an up to date approach for longitudinal modelling of %FEV<sub>1</sub> and apply it to a unique CF dataset to demonstrate its utility at the clinical and population level.

#### **Methods and Findings**

The Danish CF register contains 70,448 %FEV<sub>1</sub> measures on 479 patients seen monthly between 1969 and 2010. I partitioned the variability in the data into three components (between patient, within patient and measurement error) using the empirical variogram. I then develop a linear mixed effects model to explore factors influencing %FEV<sub>1</sub> in this population. Lung function measures are correlated for over 15 years. A baseline %FEV<sub>1</sub> value explains 63% of the variability in %FEV<sub>1</sub> at one year, 40% at three years, and about 30% at five years. The model output smooths out the short-term variability in %FEV<sub>1</sub> (SD 6.3%), aiding clinical interpretation of changes in %FEV<sub>1</sub>. At the population level I demonstrate significant effects of birth cohort, pancreatic status and *P. aeruginosa* infection status on %FEV<sub>1</sub> over time.

#### **Conclusions**

The modelling approach presented here provides a more realistic estimate of the underlying %FEV<sub>1</sub> trajectory of people with chronic lung disease, compared to the random intercept and slope approach, by acknowledging the imprecision in individual measurements and the correlation structure of repeated measurements on the same individual over time. This method has applications for clinicians in assessing prognosis and the need for treatment intensification, and for use in clinical trials.

## Introduction

Understanding the long-term natural history of changes in lung function in people with lung diseases is a research priority (Holgate, 2007). In order to do this, objective measures of disease progression are necessary. The percent predicted forced expiratory volume in one second (%FEV<sub>1</sub>) is commonly used to monitor lung function, and to describe disease severity in CF (Davies and Alton, 2009) and chronic obstructive pulmonary disease (COPD) (Rabe et al., 2007). The %FEV<sub>1</sub> is used to inform clinical decisions about changing or intensifying treatment, and as an outcome measure in clinical studies (Ramsey et al., 1993, Konstan et al., 2007a, Konstan et al., 2007b). Furthermore %FEV<sub>1</sub> has been shown to be related to survival in CF. Kerem et al's study in 1992 demonstrated that patients with a %FEV<sub>1</sub> less than 30 had a two-year mortality over 50 percent (Kerem et al., 1992), though a more recent study shows that survival rates at low levels of lung function have improved in subsequent cohorts (George et al., 2011).

Interpreting the significance of changes in %FEV<sub>1</sub> in CF to inform patient management and to counsel patients regarding prognosis requires an understanding of the inherent variability of %FEV<sub>1</sub> measures within individuals, to determine what constitutes a clinically significant deterioration in %FEV<sub>1</sub>, rather than a change due to measurement error, or recoverable day-to-day fluctuation in lung function (Hnizdo et al., 2005, Corey, 2007). Furthermore, this variability needs to be understood to make valid inferences about the association between covariates and %FEV<sub>1</sub> in observational studies.

As survival in CF improves with successive cohorts, there are many more people surviving into late adulthood. An implication of this, coupled with the availability of long term follow-up data in CF registers, is that up to date methods should be adopted to interpret the long-term dynamics of lung function in CF. Statistical techniques for longitudinal data analysis have been the subject of much methodological development over the past 20 years, and the random intercept and slope model has become a popular analysis framework (Corey et al., 1997, Harber et al., 2007, Konstan et al., 2007a, Konstan et al., 2007b, Stern et al., 2007, van Diemen et al., 2011). While this is often appropriate for relatively short follow-up periods, there are theoretical reasons to suggest that this approach will lead to incorrect

inferences if applied over longer follow-up periods. One property of the method is that the modelled variability in %FEV<sub>1</sub> increases as a quadratic function over time (in proportion to time squared), which leads to estimates that diverge unrealistically over longer time periods. Methods for undertaking these analyses over longer time periods have been described (Diggle et al., 2002), but have not been commonly applied.

Now that people are surviving with CF for much longer, how can we optimally describe how %FEV<sub>1</sub> changes over time, in a way that is useful for clinicians at both the individual and the population level? In this study I address this issue, and analyse a unique population-level dataset of people with CF that includes longitudinal %FEV<sub>1</sub> measures taken monthly for up to 30 years. I develop a general model for %FEV<sub>1</sub> decline that goes beyond the popular random-intercept and slope approach, and explicitly describes the variability in %FEV<sub>1</sub> within individuals over time. I further show how this could be applied clinically to help interpret the significance of changes in lung function, and at a population level to explore the association of covariates (e.g. *P. aeruginosa* acquisition) with %FEV<sub>1</sub> decline. In terms of the initial objectives of this thesis, being able to model lung function decline in the Danish population is a pre-requisite to exploring the effect of SES on such outcomes.

## **Methods**

### **Subjects**

All patients aged over five contributing %FEV<sub>1</sub> data in the Danish CF database between 1969 to 2010 were eligible. I excluded post-transplant data from patients who had received a lung transplant. Patients attending the two Danish CF centres (Copenhagen and Aarhus) are seen routinely every month in the outpatient clinic, for evaluation of clinical status, pulmonary function, and microbiology of lower respiratory tract secretions. It is estimated that coverage of people with CF, resident in Denmark is almost complete from 1990, when CF care was centralised. This coverage and the unparalleled frequency of measurement make this a unique dataset for epidemiological research. The Danish Data inspectorate (Datatilsynet) approved the study.

## **Lung function testing**

The primary outcome for this analysis is %FEV<sub>1</sub>. Pulmonary function tests were performed according to international recommendations (Miller et al., 2005), measuring forced expiratory volume in one second, expressed as a percentage of predicted values for sex and height using reference equations from Wang or Hankinson (Wang et al., 1993, Hankinson et al., 1999).

## **Covariates**

Covariates in the analysis were: age; sex; genotype coded as the number of delta F508 alleles (0, 1 or 2); onset of chronic *Pseudomonas* infection (coded 0 or 1 as a time-varying covariate); PI determined on the basis of pancreatic enzyme usage (coded 0 or 1 as a baseline covariate); birth cohort (six 10-year cohorts starting at 1948); and CFRD diagnosed using the WHO criteria (coded 0 or 1 as a time-varying covariate).

## **Statistical analysis**

A detailed explanation is given in the methods chapter (Chapter 3). Repeated %FEV<sub>1</sub> measures on individuals are correlated, and this must be accommodated to obtain valid inferences. I use a linear mixed-effects model with longitudinally structured correlation (Diggle et al., 2002, Fitzmaurice, 2004), and contrast this approach with the widely used random-intercept-and-slope model (Laird and Ware, 1982), which assumes that repeated measures on individual subjects vary independently around a subject-specific linear time-trend, that itself varies randomly between subjects (Laird and Ware, 1982). For data consisting of short time-sequences of observations on each subject, this model often fits reasonably well. However, for longer time-sequences it is too rigid. I therefore model random variation in %FEV<sub>1</sub> over time within an individual subject so that the strength of the correlation of the random variation between two values depends on the corresponding time-separation. The model decomposes the overall random variation in the data into three components: between subjects; between times within subjects; and measurement error.

Initially, I fit a provisional model for the mean response by OLS and used the empirical variogram of the residuals (see methods, Chapter 3) to provide initial

estimates for the three components of variation, and for the shape of the correlation function of the between-times-within-subjects component. I then re-estimated all of the model parameters by maximum likelihood, and use generalized likelihood ratio statistics to compare nested models, and Wald statistics to test hypotheses about model parameters. I assessed associations between single or multiple covariates and the population mean %FEV<sub>1</sub> over time, and explored alternatives to a linear function for the population-averaged time-trend.

To test the overall goodness-of-fit of the final model, I analysed the residuals as follows. First, plots of residuals against fitted values should show random scatter. Second, the residuals should have approximately the same covariance structure as the fitted model, which I checked by comparing their sample variogram with the theoretical variogram of the model.

All analyses were carried out using the R open-source software environment ([www.r-project.org](http://www.r-project.org)). Maximum likelihood estimation, generalized likelihood ratio tests and Wald tests used the `lme()` function within the `nlme` package, together with the exponential class of correlation functions. Variogram calculations used an R function specially written by Peter Diggle. This R code, and some illustrative code used to generate the plots, and to fit the final model, can be found in the appendix for this chapter (Appendix 6).

## Results

### Visualising the data

Figure 65 shows the Danish dataset, with six individual trajectories selected in each panel. There are clear short-term and long-term components to the overall variation in %FEV<sub>1</sub> over time. One of the key issues in this study relates to understanding the short-term dips (or improvements) in %FEV<sub>1</sub> – are these catastrophic declines, or temporary changes from which individuals can recover?

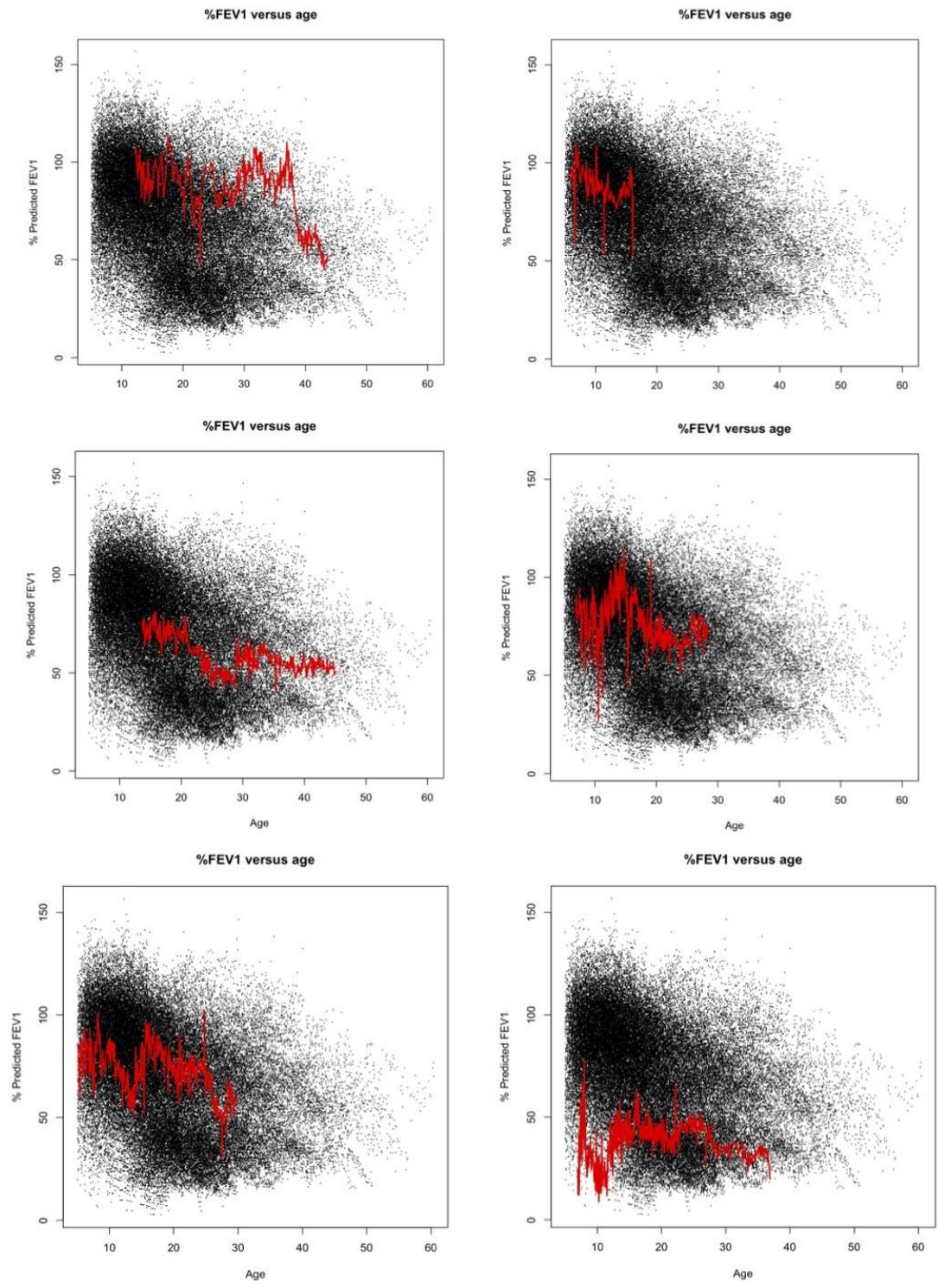
### Population characteristics

The dataset contains 70,448 lung function measures on 479 patients seen between 1969 and 2010 in Denmark (Table 17). The median number of %FEV<sub>1</sub> measures per person was 101 (range 2 to 597). The median follow-up period was 10.5 years (range 0.1 to 31.5), with a total of 6500 person-years of follow up. Forty-two patients were followed up for more than 30 years (see methods section). In the earlier birth cohorts there was selection into the dataset, with patients surviving to contribute %FEV<sub>1</sub> measures appearing in the dataset, and some patients that had died being omitted. Only 3% of the sample did not carry at least one delta F508 allele and 96% of the population had PI.

**Table 17: Baseline characteristics of Danish CF population**

	Birth Cohort						Total
	>=1948	>=1958	>=1968	>=1978	>=1988	>=1998	
N (%)	7 (1.5)	42 (8.8)	110 (23)	105 (21.9)	141 (29.4)	74 (15.4)	479 (100)
Female	1 (14.3)	19 (45.2)	48 (43.6)	52 (49.5)	74 (52.5)	42 (56.8)	236 (49.3)
No Delta F508 = 0	0 (0)	0 (0)	1 (0.9)	4 (3.8)	5 (3.5)	5 (6.8)	15 (3.1)
No Delta F508 = 1	2 (28.6)	14 (33.3)	26 (23.6)	24 (22.9)	42 (29.8)	19 (25.7)	127 (26.5)
No Delta F508 = 2	5 (71.4)	28 (66.7)	83 (75.5)	77 (73.3)	94 (66.7)	50 (67.6)	337 (70.4)
Developed chronic Pseudomonas	6 (85.7)	31 (73.8)	84 (76.4)	55 (52.4)	20 (14.2)	5 (6.8)	201 (42)
Missing infection information	0 (0)	5 (11.9)	2 (1.8)	2 (1.9)	1 (0.7)	0 (0)	10 (2.1)
PI	7 (100)	42 (100)	105 (95.5)	99 (94.3)	133 (94.3)	73 (98.6)	459 (95.8)
Copenhagen	7 (100)	38 (90.5)	83 (75.5)	72 (68.6)	79 (56)	50 (67.6)	329 (68.7)
Alive	4 (57.1)	27 (64.3)	79 (71.8)	77 (73.3)	132 (93.6)	74 (100)	393 (82)
Developed CFRD	3 (42.9)	21 (50)	41 (37.3)	31 (29.5)	22 (15.6)	1 (1.4)	119 (24.8)

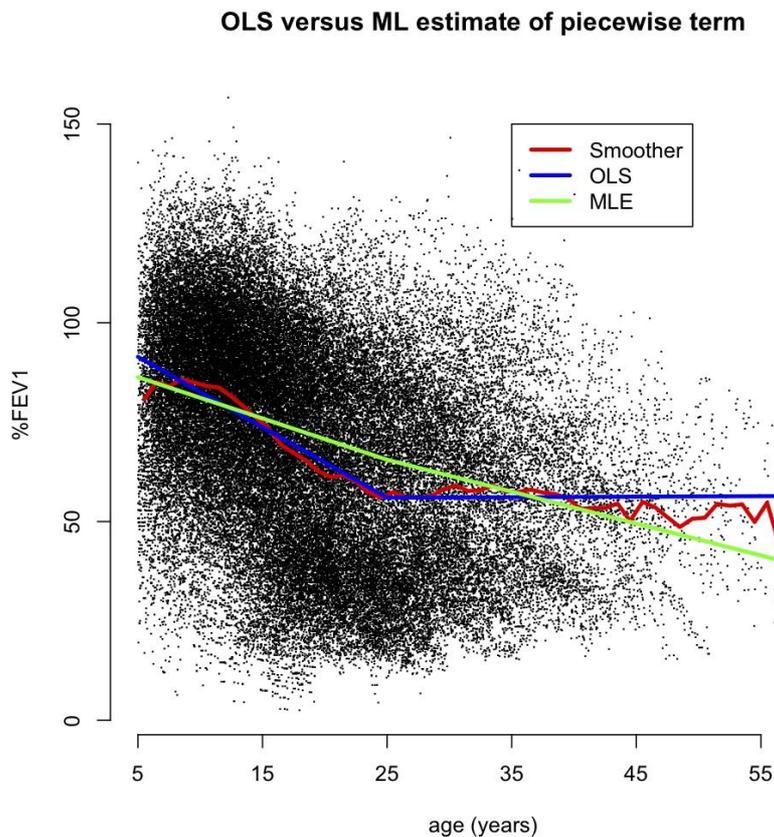
**Figure 65: Scatterplot of %FEV<sub>1</sub> versus age in Denmark, with a randomly selected individual highlighted in each panel**



## Exploring the form for the population average

Figure 66 shows all the %FEV<sub>1</sub> measures over time in a scatterplot, with an added mean smoother in red. The mean smoother does not take into account the correlation of repeated measures within individuals. There appears to be a fairly linear decline in mean %FEV<sub>1</sub> to about age 25, where the mean stabilizes, before becoming more erratic at older ages where the sample size is smaller. A piecewise OLS regression (blue line) with a change in slope at age 25 provides an improved fit over a straight line, and by eye one can see that this fits the smoother well. However, when one fits the same piecewise mean in the longitudinal model using ML estimation, then the change in slope at age 25 is not significant (green line). This indicates that the levelling off of the mean smoother to some extent reflects selective drop out (death) (see methods section). The final longitudinal model implicitly takes this drop out into account and generates the parameter estimates that one would expect to see if dropout had not occurred. I therefore modelled the population average as a straight line.

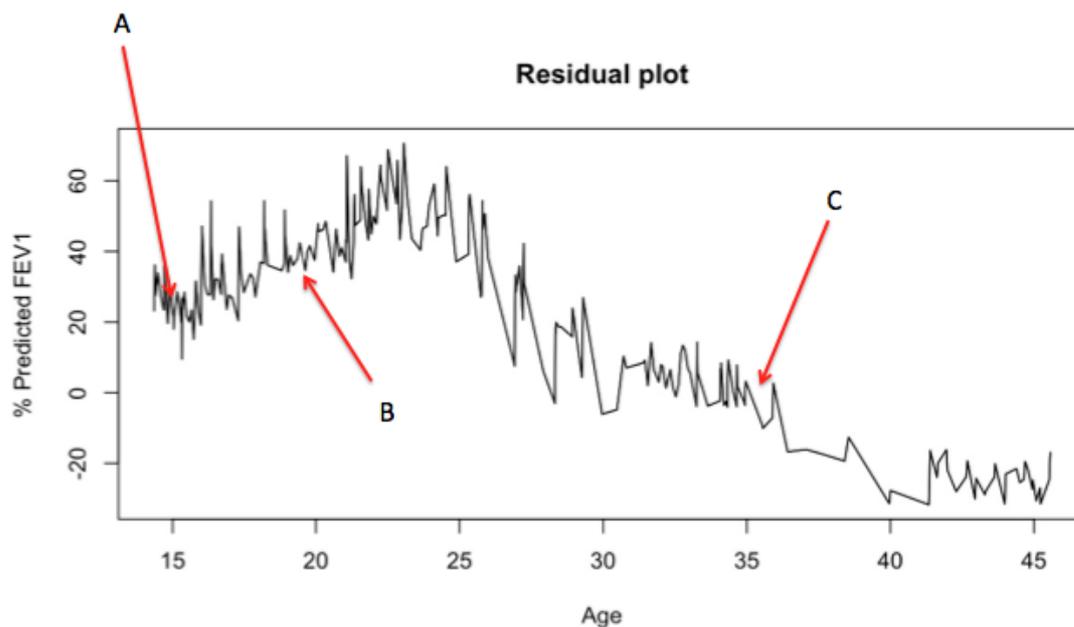
**Figure 66: All Danish data with smoothed mean trend**



### Quantifying the variability in %FEV<sub>1</sub> over time

Full technical details of the empirical variogram of the residuals are provided in the methods section (Chapter 3). A residual trajectory is shown in Figure 67 below. The empirical variogram function summarises the average variance within individuals in the population at successive time lags. For instance for the individual in Figure 67 one can calculate the variance of the difference between %FEV<sub>1</sub> comparing points A and B, at a time lag of five years, and likewise the variance of the difference at points A and C at a time lag of 20 years. Repeating this for all time lags in a particular individual, and pooling the results over the population yields the empirical variogram.

**Figure 67: Residual plot for individual in Danish population**

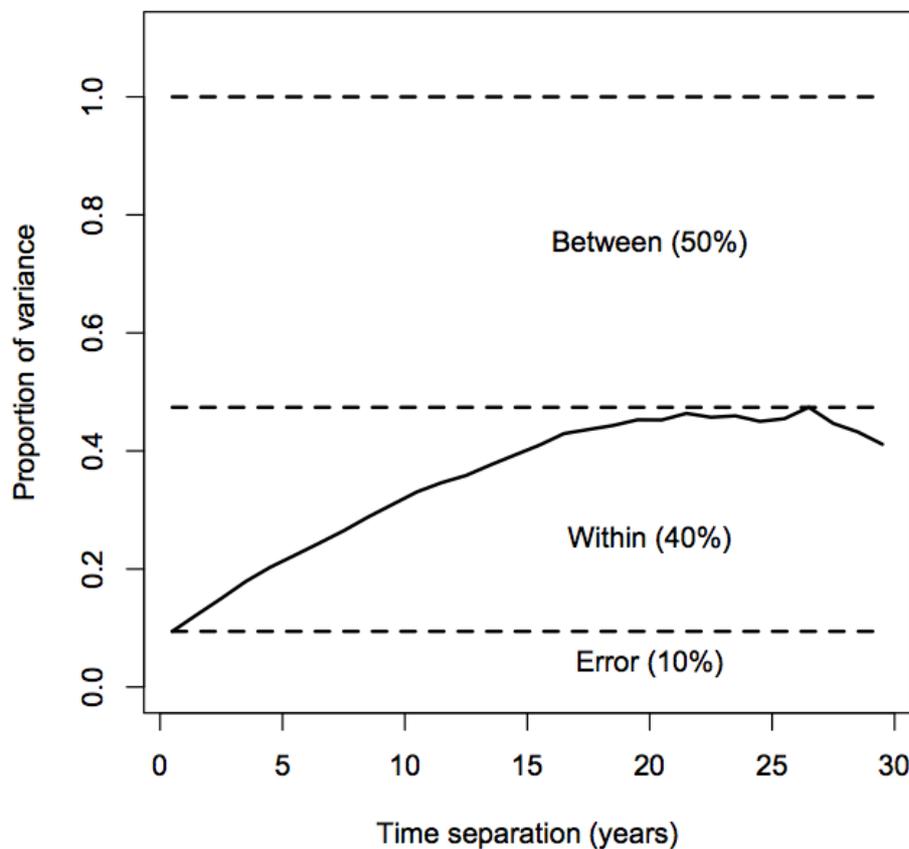


The empirical variogram estimates the three components of variability in the dataset (error, within individual over time, between individual) (Figure 68). The intercept at time zero represents measurement error, since there can be no true within-person variation at a time lag of zero. Of the total variance in the Danish dataset, about half is due to systematic differences between patients (e.g. genotype, sex or pancreatic status), two-fifths is within patients, representing change over time (disease progression), and one tenth is 'measurement error'. In practice, this last component represents the combined effects of technical errors, and physiological variability occurring at time intervals less than the monthly interval of measurement, e.g. day to

day variability. This error variance equates to an average standard deviation of 6.3% for repeated measures on the same individual at short time-intervals.

**Figure 68: Quantifying the variability in %FEV<sub>1</sub> with the variogram approach**

*The figure shows the scaled empirical variogram for the Danish data. The solid line (variogram function) represents the variance of the difference between residual errors within individuals at time lags from zero to 30 years. The variogram function increases up to about 15 years, corresponding to a decreasing correlation between paired lung function measures with increasing time-separation. The variogram partitions the variability in the data into three components: within person, between person, and error.*



Hence, the sample variogram provides initial estimates of the variance components  $\tau^2$ ,  $v^2$  and  $\sigma^2$ , and of the correlation function  $\rho(u)$  (see methods for further details). An alternative representation is to transform  $V(u)$  into a function  $R^2(u)$ , defined as

$$R^2(u) = (1 - V(u)/\sigma^2)^2$$

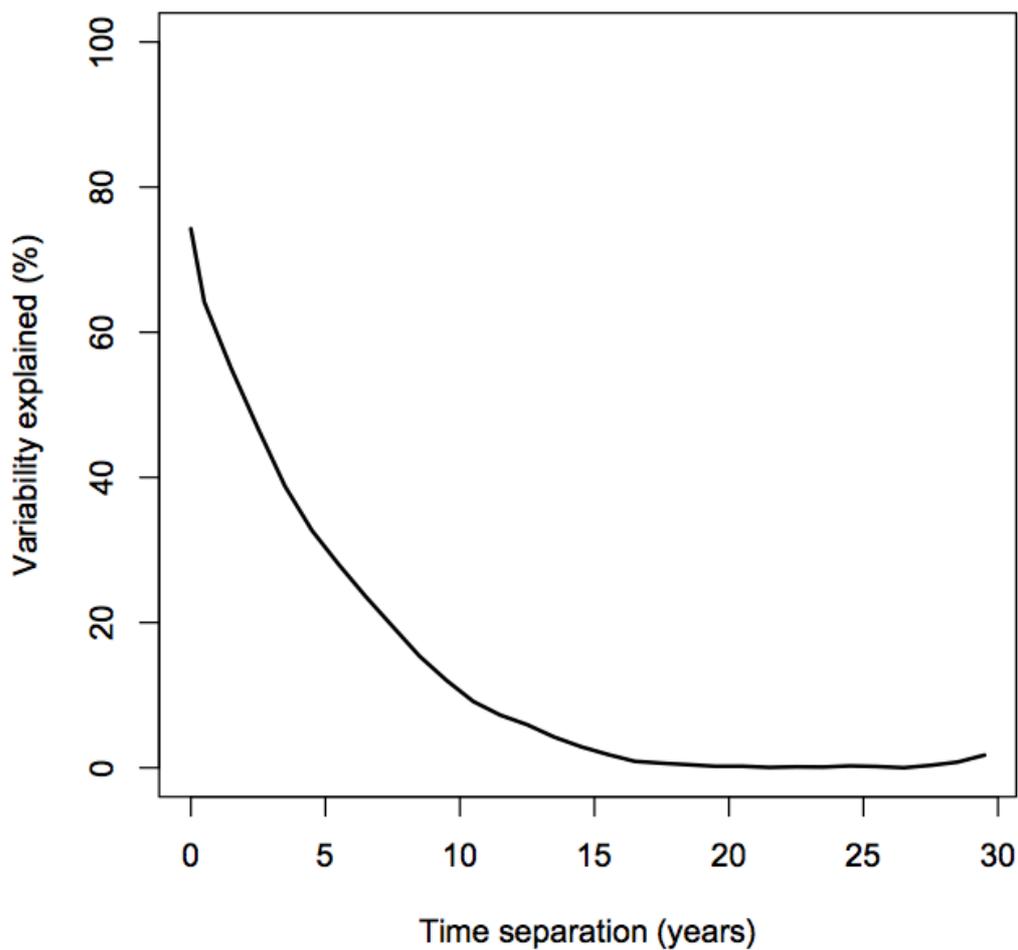
This function is analogous to the conventional  $R^2$ -value for a fitted regression model in the sense that it measures the proportion of within-patient variation in a person's

lung-function at time  $t+u$  that can be explained by their lung-function at time  $t$  (Figure 69).

Figure 69 shows the proportion of the within-person variability in %FEV<sub>1</sub> at follow-up time ( $t$ ) that can be explained by their %FEV<sub>1</sub> value at baseline. For example, about 50% of the within-patient variability at  $t = 2.5$  years is explained by the baseline measurement, and about 30% at  $t = 5$  years. Overall, the dependence on baseline measures gradually decays and is negligible at 15 years.

**Figure 69: Proportion of variability in an individual's %FEV<sub>1</sub> at follow-up time  $t$  that is explained by their %FEV<sub>1</sub> at baseline.**

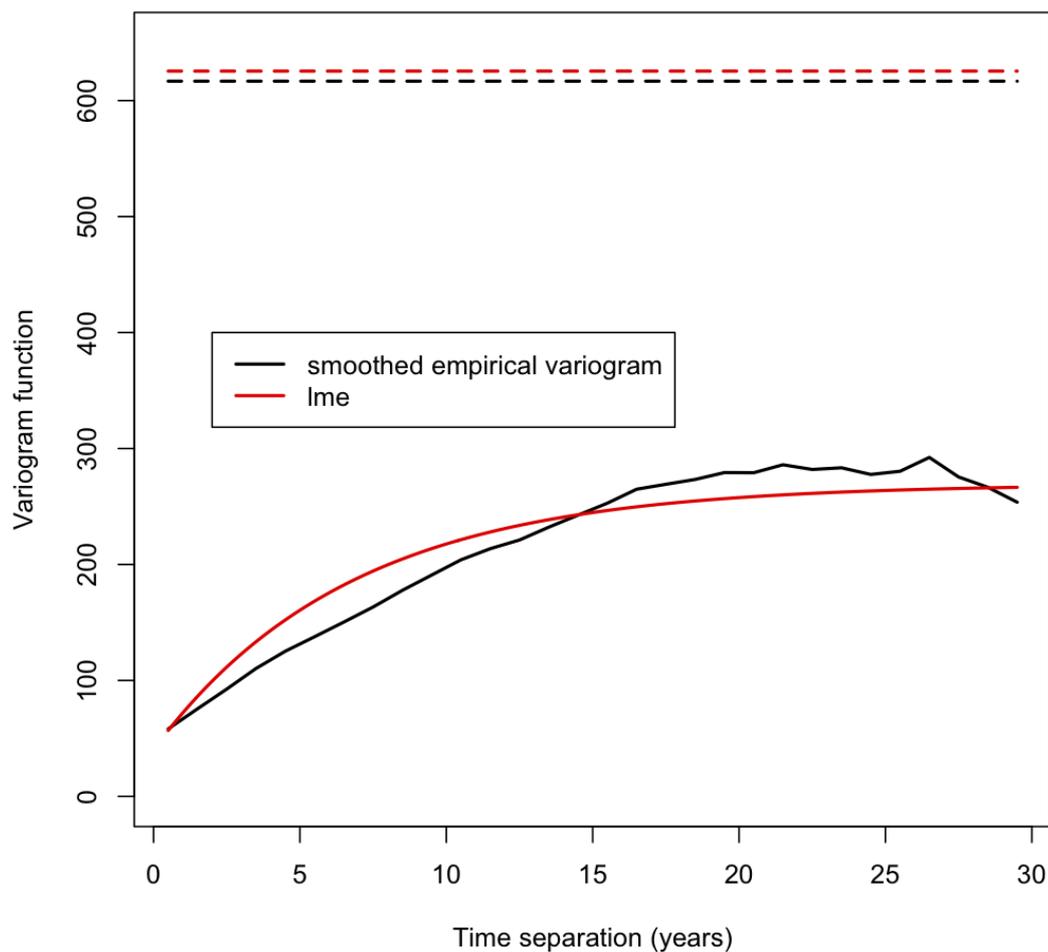
*This shows that the variogram can predict 63% of the variability from the population average at one year, which decreases to around 60%, 40%, 30% and 10% at 2,3,5 and 10 years respectively.*



### Fitting the final model

Figure 70 below compares the empirical variogram fit to the theoretical variogram plotted using the MLE estimates from the lme() function in R, with an exponential correlation. This shows that the modelled correlation function approximates reasonably to the empirical correlation in the dataset.

**Figure 70: Comparison between empirical variogram and MLE variogram estimate**

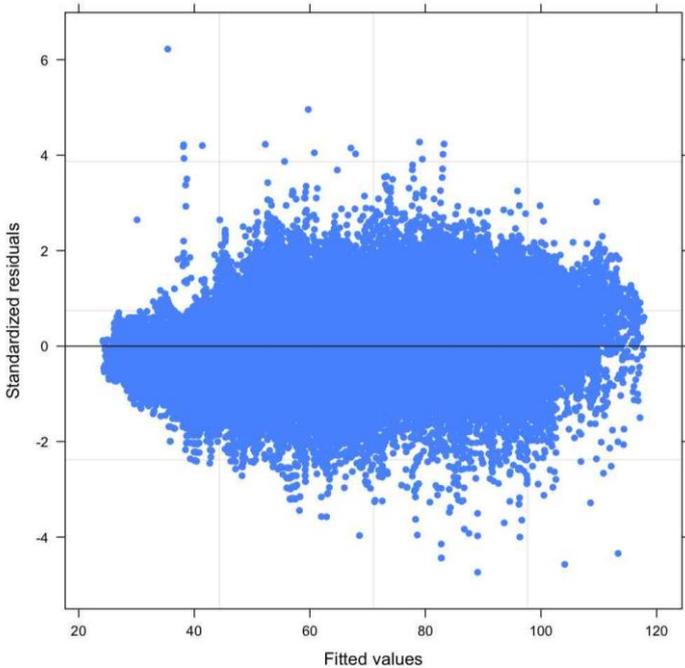


The estimated variance ( $\%FEV_1^2$ ) components derived from the modelled variogram are as follows: total variance=625, within-person variance=229, between-person variance=356, error variance=40.

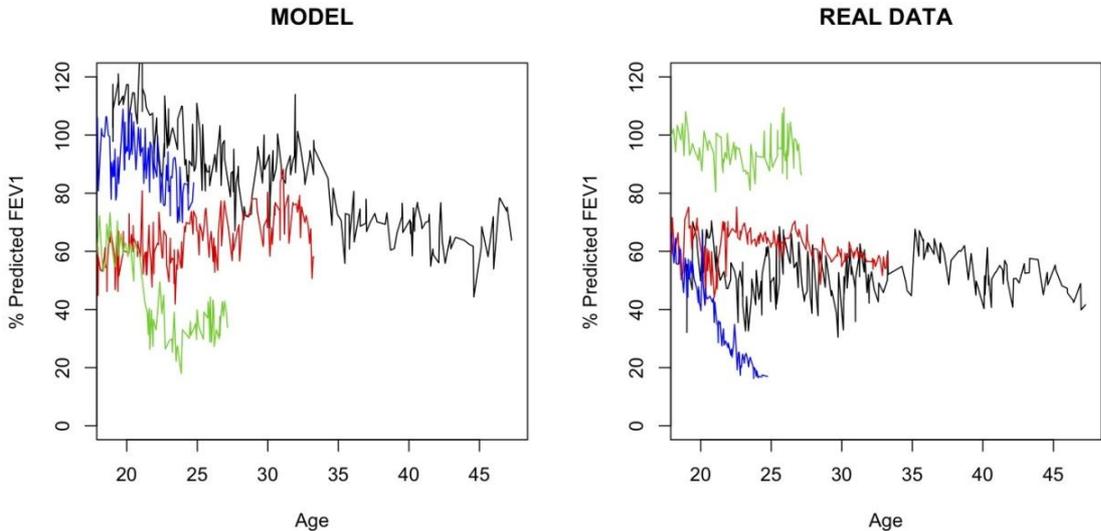
**Assessing model fit**

Figure 71 below plots the standardised residuals against the fitted values. There are no trends in the residuals, and there is no evidence of non-constant variance. As a further check, I simulated trajectories from the fitted model, and compared them to raw data traces (Figure 72). This demonstrates that the model is generating individual level predictions that have a similar form to the real data.

**Figure 71: Scatterplot of standardised residuals versus fitted values**



**Figure 72: Simulated realisations from the final model**



### **Comparing the stationary Gaussian process (SGP) model to the random intercept and slope approach**

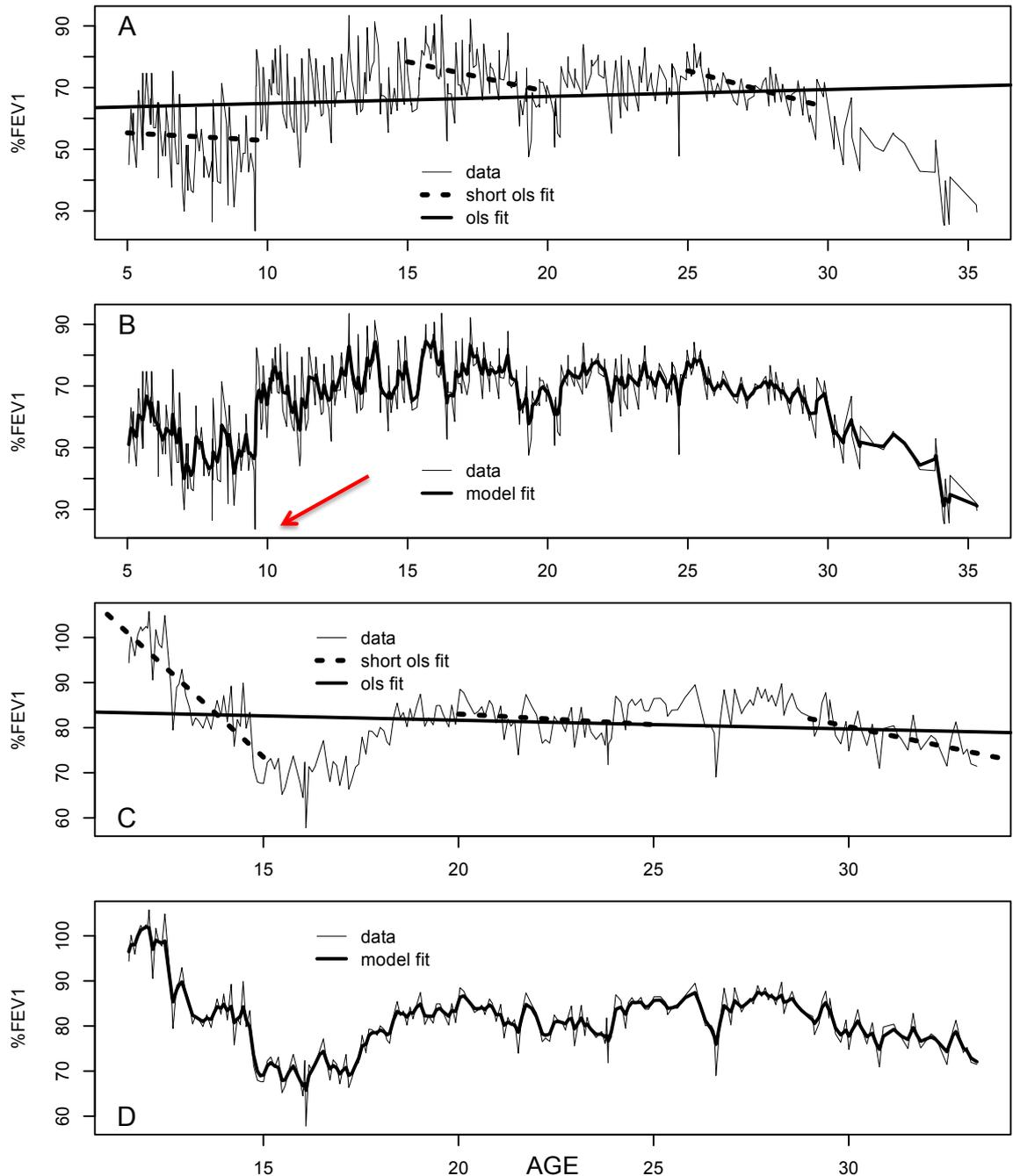
The high degree of short-term and long-term variation in predicted %FEV<sub>1</sub> are illustrated in Figure 73. The standard random intercept and slope model approach is illustrated over long and short follow up periods in Figure 73 A and C. This approach assumes that any deviation of an individual's trajectory from the population mean is linear in time over the whole of the follow-up period apart from independent random errors. One can see that this assumption is reasonable over short time-periods, as illustrated by the fit of the shorter dotted-line segments (Figure 73 A and C), but over longer time-periods the individual data traces diverge unrealistically from their fitted linear mean trajectories (long solid line). The proposed stationary Gaussian process (SGP) model produces a much closer fit to the data (Figure 73 B and D), and one that better reflects the relative magnitude of the three estimated components of variation in %FEV<sub>1</sub> over time. The smoothed fitted trace is a better representation of the 'true' underlying lung function, and could be used in real time to guide the interpretation of sudden changes in lung function. For instance, the sudden drop to under 30% indicated by the arrow is not mirrored in the model trace, suggesting that this may be recoverable random fluctuation.

### **Clinical utility of the proposed model**

The model can be used to guide interpretation of sudden changes in lung function. Consider seeing the person in Figure 73 B at around age nine (as indicated by the arrow in the figure), when her lung function has dropped to below 30%. On the basis of this one-off measurement, one might be quite guarded in terms of prognosis. However, our modelled trace (thick black line in Figure 73 B) suggests that her underlying lung function is changing less dramatically, with a modelled %FEV<sub>1</sub> of around 50%. I suggest that this estimate provides a more realistic assessment of underlying lung function by smoothing out the short-term variability. This could be a useful adjunct to clinical decision-making. As well as providing information about the significance of a sudden change in lung function, Figure 69 also quantifies the predictive value of a contemporary %FEV<sub>1</sub> measure. In terms of counselling patients, this means that a higher %FEV<sub>1</sub> today is associated with a higher %FEV<sub>1</sub> at subsequent time points, but the predictive value deteriorates over time as illustrated in the figure.

**Figure 73: Comparison of conventional random-intercept and slope model over short and long follow-up periods, versus proposed Gaussian process model.**

Panel A shows the data for a single individual, illustrating that a linear trend fits reasonably well over short time-periods, but gives a very poor fit to this individual's complete data; linear trends are fitted by OLS. Panel B shows the same data with the fitted trajectory of the stationary Gaussian process model. Panels C and D show the corresponding plots for a second individual.



## Effect of covariates on lung function in the Danish population

I explored the effect of covariates that have been associated with %FEV<sub>1</sub> in previous studies, in order to demonstrate how this model can be used to answer questions at the population level, as per Konstan et al's study (Konstan et al., 2007a). I first explored univariate associations, which are shown in Table 18. Figure 74 demonstrates the effect of pancreatic status in the univariate analysis.

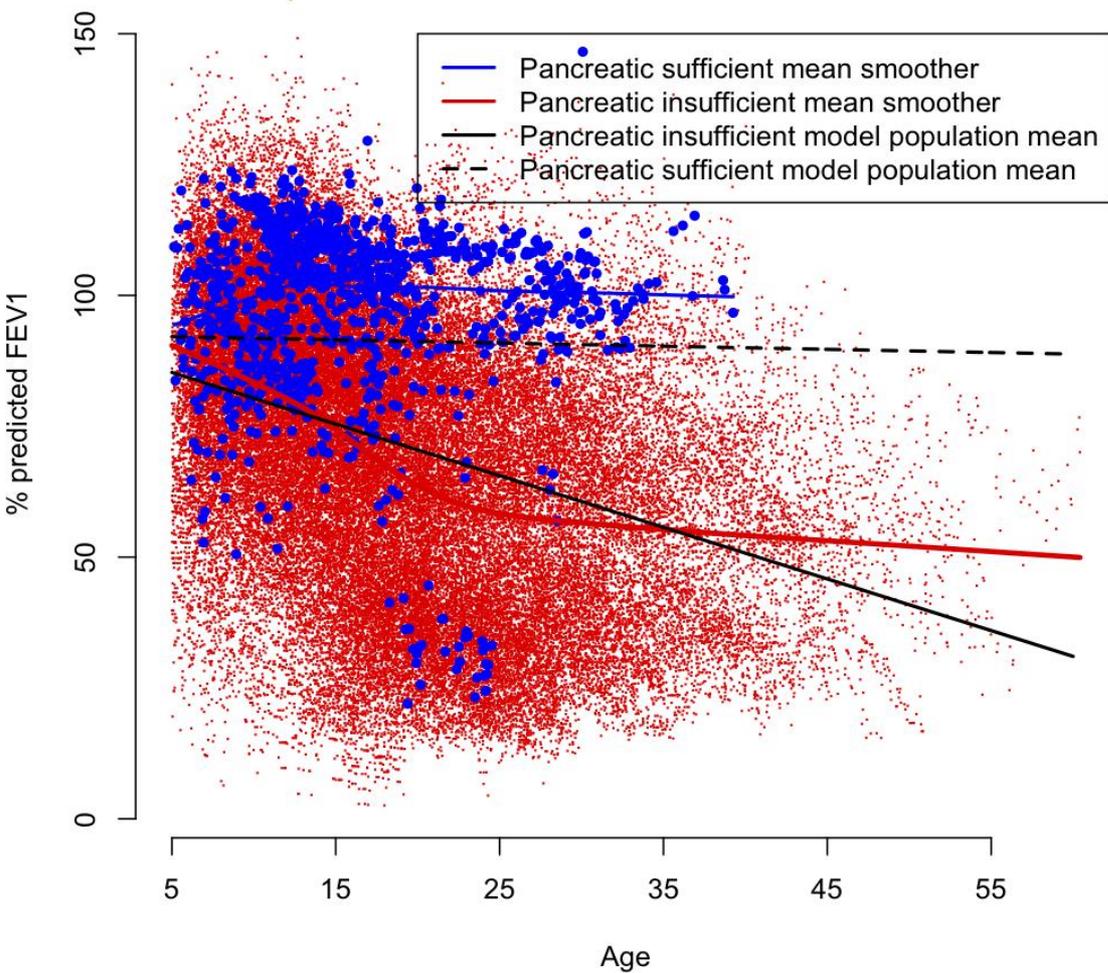
The final model included age, *Pseudomonas* status, pancreatic status, cohort and CFRD (Table 19). Note that the estimated covariate effects in Table 19 are population-averaged effects, i.e. they describe average values of %FEV<sub>1</sub> for sub-populations of individuals sharing the same explanatory characteristics, rather than for any one individual. The most prominent effects are associated with birth cohort, pancreatic function and the onset of *Pseudomonas* infection (Figure 75). There is clear separation between the three most recent birth cohorts, with a successive increase in the intercept term at age 5 (83% in the 1978 to 1988 cohort versus 96% in the post-1998 cohort) (Figure 75, Figure 76, Figure 77). There is a large change in the point estimate for the rate of change of lung function in the post-1998 (0.24%) compared with the 1988-98 cohort (-1% per year), such that the post-1998 cohort appears to be improving over the period of measurement. The three cohorts spanning the years 1948 to 1978 have a similar overall rate of decline around -0.3% per year, with an intercept at age five of 66%. PI is associated with a significantly steeper rate of decline of lung function (-0.92% per year 95%CI -1.7 to -0.3), as is acquisition of *Pseudomonas* infection (-0.5% per year 95%CI -0.72 to -0.3) (Figure 75). CFRD is associated with a drop in intercept of -2.5% (95% CI -3.6 to -1.37), but has no effect on the rate of decline of lung function.

**Table 18: Univariate associations between covariates and %FEV<sub>1</sub>**

	Value (%FEV <sub>1</sub> )	Std.Error	p-value
<b>Genotype</b>			
(Intercept)	90.530	6.656	<0.0001
age (years)	-0.785	0.454	0.0837
Number delta F508=1	-2.801	7.075	0.6923
Number delta F508=2	-5.758	6.803	0.3978
age:Number delta F508=1	-0.070	0.466	0.8805
age:Number delta F508=2	-0.219	0.458	0.6327
<b>Pancreatic insufficiency</b>			
(Intercept)	92.101	7.144	<0.0001
age (years)	-0.060	0.362	0.868
Pancreatic insufficiency	-6.822	7.244	0.3468
age:Pancreatic insufficiency	-0.926	0.366	0.0114
<b>Sex</b>			
(Intercept)	84.907	1.680	<0.0001
age (years)	-1.015	0.077	<0.0001
Male	1.747	2.393	0.4657
age:Male	0.091	0.105	0.3868
<b>Cohort</b>			
(Intercept)	70.278	2.375	<0.0001
age <sup>2</sup>	-0.640	0.084	<0.0001
cohort>=1948 (reference 1968)	9.082	13.988	0.5165
cohort>=1958	1.252	4.860	0.7968
cohort>=1978	15.062	3.387	<0.0001
cohort>=1988	22.492	3.159	<0.0001
cohort>=1998	26.813	3.640	<0.0001
age:cohort>=1948	-0.149	0.337	0.6571
age:cohort>=1958	-0.024	0.149	0.8717
age:cohort>=1978	-0.538	0.142	0.0001
age:cohort>=1988	-0.356	0.178	0.0459
age:cohort>=1998	0.884	0.464	0.057
<b>Pseudomonas</b>			
(Intercept)	84.714	1.224	<0.0001
age (years)	-0.738	0.081	<0.0001
onset of Pseudomonas	-0.392	0.104	0.0002
<b>CFRD</b>			
(Intercept)	85.670	1.198	<0.0001
age (years)	-0.928	0.054	<0.0001
CFRD	-2.861	1.219	0.019
age:CFRD	0.024	0.064	0.701

**Figure 74: Univariate effect of pancreatic status**

*Raw data disaggregated by pancreatic status, with pancreatic sufficient individuals coloured in blue.*

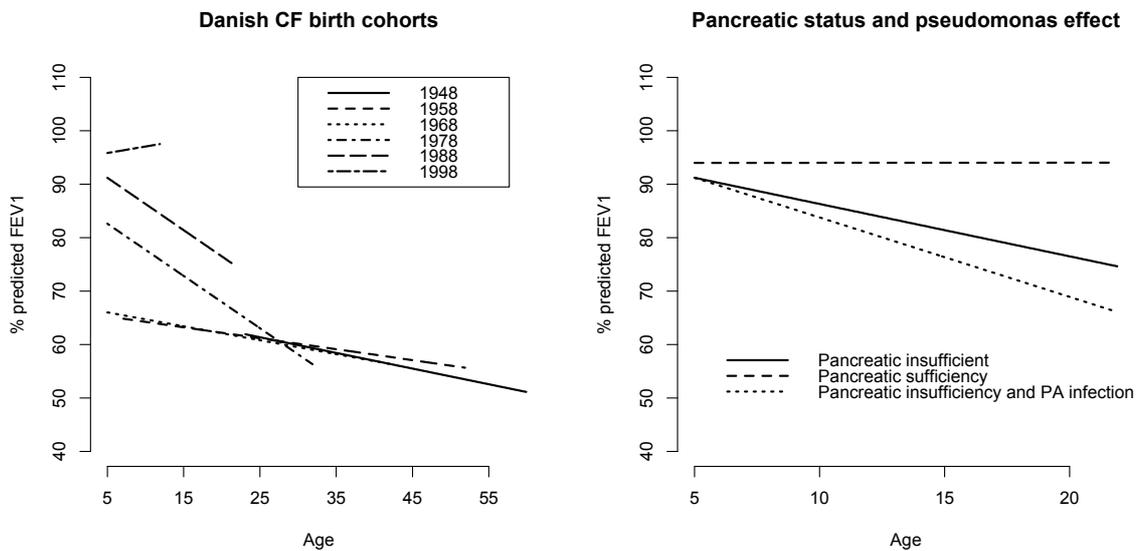


**Table 19: Estimates from final multivariate model**

	Point estimate	lower 95% CI	upper 95% CI	p-value
Intercept at age 5 years	66.02	61.13	70.92	<0.001
CFRD	-2.47	-3.58	-1.37	<0.001
age	-0.26	-0.49	-0.03	0.025
cohort $\geq$ 1948 (reference 1968)	1.20	-25.50	27.90	0.930
cohort $\geq$ 1958	-0.75	-10.01	8.51	0.874
cohort $\geq$ 1978	16.60	10.15	23.05	<0.001
cohort $\geq$ 1988	25.19	19.11	31.27	<0.001
cohort $\geq$ 1998	29.81	22.85	36.78	<0.001
Pancreatic sufficiency	2.78	-10.43	15.99	0.679
<i>P. aeruginosa</i> infection	-0.51	-0.72	-0.29	<0.001
age $\times$ cohort $\geq$ 1948	-0.03	-0.67	0.61	0.920
age $\times$ cohort $\geq$ 1958	0.06	-0.23	0.34	0.699
age $\times$ cohort $\geq$ 1978	-0.72	-1.00	-0.44	<0.001
age $\times$ cohort $\geq$ 1988	-0.72	-1.09	-0.35	<0.001
age $\times$ cohort $\geq$ 1998	0.50	-0.41	1.42	0.280
age $\times$ Pancreatic sufficiency	0.98	0.29	1.67	0.005

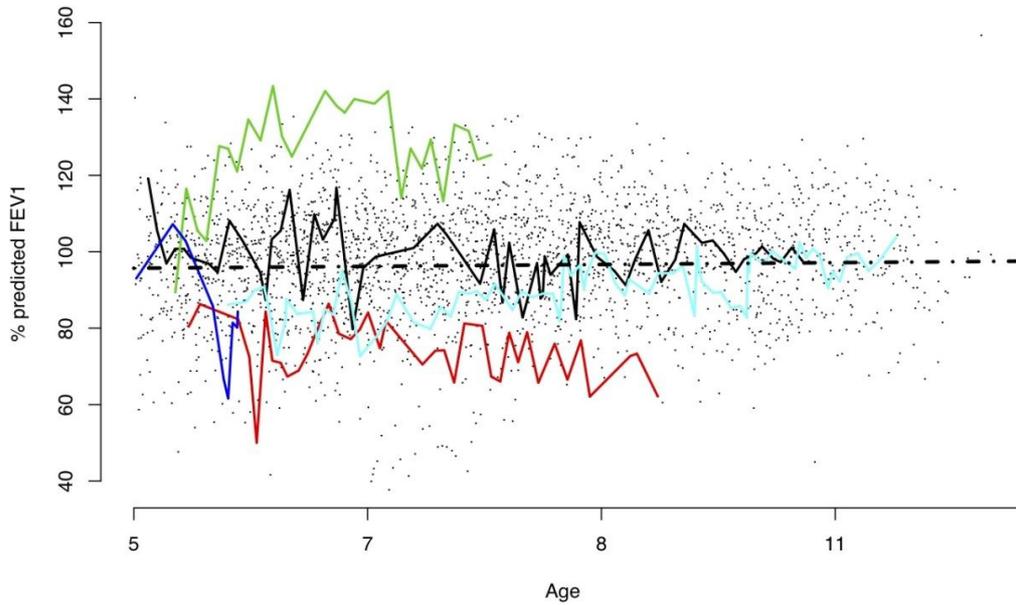
**Figure 75: Effect of key covariates on %FEV<sub>1</sub>**

The left panel shows the birth cohort effect in the final model. There is clear separation between the three most recent birth cohorts, with a successive increase in the intercept term at age five. The right panel illustrates the effect of PI and *Pseudomonas* infection on the predicted population trajectory for a person born in the 1988-98 cohort.



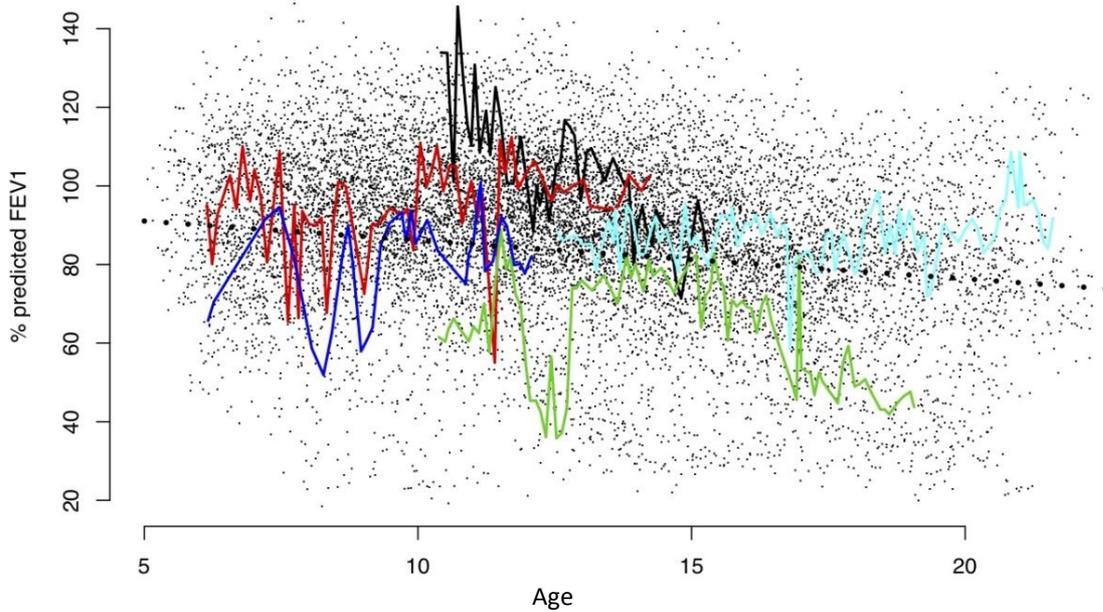
**Figure 76: Scatterplot of data for post-1998 cohort**

*Five individual traces randomly picked out, and linear population estimate from final model superimposed*



**Figure 77: Scatterplot of data for 1988-1998 cohort**

*Five individual traces randomly picked out, and linear population estimate from final model superimposed*



## Discussion

In this study I describe a novel longitudinal modelling technique specifically aimed at analysing long sequences of repeated measurements, and apply this to %FEV<sub>1</sub> from a CF population. I show how this approach could be used to inform patient management, by aiding the interpretation of sudden changes in lung function, and by quantifying the predictive value of a baseline %FEV<sub>1</sub> measure up to 15 years later. At the population level, I show how the model can be used to quantify the effect of covariates on populations or subpopulations. Translation of these methods into clinical practice is important, since people with CF are surviving for longer periods, and I have shown how commonly applied approaches are unhelpful over long follow-up periods.

This study quantifies the short-term variability in %FEV<sub>1</sub> in this population (SD 6.3%), and demonstrates that %FEV<sub>1</sub> measures within individuals are correlated over time-lags of 15 years or more. I have also explored the effect of previously studied risk factors for lung function decline in the Danish CF population, and have demonstrated significant effects of birth cohort, pancreatic status and *Pseudomonas* infection status.

## Clinical implications

The findings from this study have a number of clinical implications. Quantifying the variability in lung function measures is essential in order to make correct clinical interpretation (Hnizdo et al., 2005). Exploiting the unusually high frequency of data collection in Denmark, this study implies that on average a change in %FEV<sub>1</sub> of greater than 13% (i.e. twice the error SD, to give a 95% confidence range) is likely to represent true within-patient variation over time (disease progression), whereas anything less than this could be due to short-term fluctuation, which may recover. Stanbrook et al (Stanbrook et al., 2004) found a pooled within-subject %FEV<sub>1</sub> SD of 4.5% when measured over a nine-day period in 21 stable adults with CF. This population is different to the population in this study, who were measured regardless of clinical status, and one would therefore expect greater variability. Other studies have shown that people with CF, asthma and COPD have more short-term variability in lung function tests (Hruby and Butler, 1975, Pennock et al., 1981, Cooper et al.,

1990) and that more impaired lung function is associated with greater variability (Enright et al., 2004).

The presence of the significant measurement error described in this study presents a challenge to the use of isolated %FEV<sub>1</sub> measures for long term prediction of prognosis. A number of methods have been proposed for predicting future changes in %FEV<sub>1</sub> to inform clinical decision-making. Vandevanter et al have developed a short term pulmonary outcome prediction tool, for estimating the change in lung function over a 4-year period using cross sectional regression models (Vandevanter et al., 2010). Konstan et al use random-intercept and slope models over short follow-up periods to develop a regression equation that estimates the effect of baseline covariates on subsequent lung function decline (Konstan et al., 2007a).

At the population level we show how this approach can be applied to quantify the effect of covariates on changes in lung function. Furthermore, the partitioning of the variability in %FEV<sub>1</sub> and the precise description of the correlation structure captured in the model provide important information for sample size calculations in longitudinal clinical studies with %FEV<sub>1</sub> as an outcome. Increasingly longitudinal outcomes are being used in RCTs, and in order to undertake an a priori sample size calculation it is essential to have information on the correlation structure. Furthermore, the modelled %FEV<sub>1</sub> trace could be used as an outcome in its own right.

The modelling approach can be used to generate an underlying representation of an individual's 'true' lung function trajectory (Figure 73) that smoothes out the noise inherent in %FEV<sub>1</sub> measures. These smoothed traces could be used to inform clinical decision-making – the model fit curves in Figure 73 provide more realistic estimates of underlying lung function, and more valid criteria for clinical decisions. I propose that this model could be used to develop a real-time smoothing tool embedded in electronic patient records to aid clinical interpretation of spirometry data. I further suggest that access to this information would provide some re-assurance to patients experiencing lower than expected lung function values, since lung function can recover quite dramatically, and these data suggest that a linear or stepwise decline in lung function over time is not the norm.

I have generated, for the first time to my knowledge, the variogram function for %FEV<sub>1</sub> in people with CF over long follow-up periods. This precisely quantifies how %FEV<sub>1</sub> measures are correlated over time. Furthermore, I have done this for the whole CF population of Denmark. This quantifies the degree to which a baseline %FEV<sub>1</sub> measure can be used to predict subsequent %FEV<sub>1</sub> measures over long follow-up periods, and is likely to be of interest to clinicians and patients. The analysis demonstrates a long-term correlation between levels of %FEV<sub>1</sub> within an individual. This suggests that there is long-term predictive value in a high %FEV<sub>1</sub> measure – people with CF with a high %FEV<sub>1</sub> at baseline are more likely to have a high %FEV<sub>1</sub> up to 15 years later than individuals with a lower baseline %FEV<sub>1</sub> (Figure 69). However, the predictive value of a %FEV<sub>1</sub> measure drops away rapidly over this period. We can say that on average an %FEV<sub>1</sub> reading today explains about 63% of the variability in %FEV<sub>1</sub> at one year, 40% at three years, and about 30% at five years.

This corroborates Rosenthal's study (Rosenthal, 2008), which found that baseline %FEV<sub>1</sub> explains 66% of the variability in %FEV<sub>1</sub> one year hence and Mastella et al's study of European registry data where differences in lung function at enrolment at age five, categorised as mild, moderate, or severe, tracked through the study to age 40 (Mastella et al., 2000). Konstan et al also describe how a lower %FEV<sub>1</sub> for a given age can be used to characterise the aggressiveness of lung disease (Konstan et al., 2009). Other studies have shown a high %FEV<sub>1</sub> to be an independent risk factor for a greater rate of decline of %FEV<sub>1</sub> over the next few years (Konstan et al., 2007a) (Vandevanter et al., 2010). This is not at odds with the findings here; a high %FEV<sub>1</sub> can be a risk factor for greater decline in the short term, whilst still being associated with a relatively higher %FEV<sub>1</sub> over the longer term (Konstan et al., 2009).

### **Novel modelling approach**

The approach to modelling changes in %FEV<sub>1</sub> presented here can be applied over long follow-up periods. This is in contrast to the widely used random intercept and slope approach that has been applied in studies of CF and COPD, over short-term (Vestbo et al., 1999, Mastella et al., 2000, Konstan et al., 2007a, MacLean et al., 2008) and longer-term follow up periods (Corey et al., 1997, Hnizdo et al., 2005, Stern et al., 2007, Kohansal et al., 2009). That approach has also been used to

quantify lung function decline in other diseases, such as sickle cell disease (MacLean et al., 2008) and bronchiectasis (Twiss et al., 2006). In a recent study McKay et al propose a novel approach to modelling lung function decline in an adult population (McKay et al., 2011), involving a linear mixed-effect model with a cubic spline to account for non-linear population-averaged decline in lung function with age and a standard random intercept and slope model to account for within-patient variability. The spline gives a more flexible set of models for the population-averaged trajectory, but our analysis of the Danish dataset shows that the random intercept and slope model is too rigid to capture the pattern of within-patient variability in %FEV<sub>1</sub> over longer periods. The approach presented here can easily be combined with the spline model to describe non-linear population-averaged decline, since the form of the population average and the within-patient correlation structure are separate issues, but as illustrated above this was not needed for the Danish dataset. Note also that an adequate model for the correlation structure in the data is necessary for prediction at the individual patient level.

### **Effect of covariates**

As with other studies of CF patients (Elborn et al., 1991), there is a striking cohort effect evident in this population. The treatment of CF lung disease has been transformed over the period captured in this analysis, from 1969 to the present day. In Denmark a regimen of elective IV antibiotics for 14 days every third month was introduced in 1976 for patients chronically infected with *P. aeruginosa*. Enteric-coated pancreatic enzymes were introduced in the 1980s, and in the last two decades we have seen the introduction of inhaled therapies (bronchodilators, steroids and antibiotics), and DNase. Particularly impressive is the improvement in lung function in the post-1998 cohort by comparison with preceding birth cohorts. Although this group are early in their disease progression, the overall picture suggests that new therapeutic strategies are continuing to provide improvements in respiratory function in CF.

Pancreatic sufficiency had an important effect on the overall rate of decline of lung function (+0.9 % per year). In Konstan's study (Konstan et al., 2007a) pancreatic sufficiency was the most important protective factor in the 6 to 8 year-old age group (+1.33% per year). The small number of pancreatic-sufficient individuals in the

Danish dataset (n = 20, 5%) have a notably different lung function phenotype, maintaining near-normal lung function over the period of follow-up (Figure 74). The onset of *Pseudomonas* infection was associated with a significant increase in the rate of decline of lung function, by around -0.5% per year, similar to that reported by Konstan where *Pseudomonas* colonization was associated with an increased rate of decline of FEV<sub>1</sub> of -0.31% per year in the 6 to 8 age group, and -0.22 in the 9 to 12 age group (Konstan et al., 2007a). This highlights the importance of firstly delaying the onset of chronic infection, and aggressive treatment subsequently.

### **Strengths and limitations**

The development and testing of the new approach is facilitated by the nature of the Danish CF register – to my knowledge there are no other datasets that contain such frequent (monthly) measures of lung function on individuals measured over very long periods (up to 31.5 years). CF care in Denmark is free at the point of access, and delivered from two centres in Copenhagen and Aarhus. The Danish CF Centre was originally established in 1968 in Copenhagen, and all patients connected with the centre have been seen monthly in the outpatient clinic. A similar pattern of care is undertaken in the Aarhus centre, which was established in 2000.

However, the fact that the data are from Denmark does not influence the validity of the methods we have described, since these are essentially context-free. Furthermore, this method does not exploit any features of the data that are unique to CF, and is equally applicable to other clinical areas that generate long sequences of repeated measurements. As a next step I recommend applying this method to longitudinal data collected in other CF registries, such as the UK, to clarify how robust this approach is in terms of predicting changes in %FEV<sub>1</sub> over time, and to better understand how this might inform clinical decision making. Future research could explore the utility of the proposed model in other diseases such as COPD.

A limitation of this study is the likely influence of survivor bias on lung function estimates in the earlier birth cohorts. In the 1948 to 1978 period, the intercept at age five appears significantly lower than in the other cohorts, but there is also a shallower rate of decline of lung function. This is likely to be due to the incomplete capture of patients in earlier cohorts, with censoring due to death leaving only the

more stable survivors. This is a common problem in datasets of this type (Frederiksen et al., 1996). Fitting the model by maximum likelihood automatically corrects for selection bias that depends on a patient's observed lung function measurements prior to death, although not for any additional dependence on unmeasured features of their lung function trajectory (Diggle et al., 2002, Fitzmaurice, 2004).

### **Interpretation**

The modelling approach presented in this study provides a more realistic estimate of the %FEV<sub>1</sub> trajectory in CF, which clinicians could apply in real-time to help interpret the significance of changes in %FEV<sub>1</sub>. Furthermore, this approach quantifies the predictive value of a baseline %FEV<sub>1</sub> measure, over three decades. This method is equally applicable to the longitudinal assessment of %FEV<sub>1</sub> in other lung diseases, and can enable more robust comparisons of populations, including groups studied in clinical trials. As people are now living for many decades with these diseases, the development of tools better to understand the natural history of this important outcome will be essential for better clinical care, as well as being a key research priority (Holgate, 2007).

Analysis of the Danish dataset from an equity perspective is on-going. This study has outlined the methods for analysing this unique dataset, and the next step is to use this approach in conjunction with data linked through the Danish social registers. At the time of writing, I have been able to undertake a preliminary analysis of the effect of parental education level on lung function decline in the Danish population, and this is presented briefly in Chapter 7, in the 'Research currently in progress' section.

## Chapter 7: Discussion

### Introduction

This chapter begins by summarising the key findings in relation to the objectives 1-3 of this thesis. I then describe the contribution of the studies to the existing literature, which extends beyond the initial objectives, in three main areas: CF epidemiology; our understanding of health inequalities more generally; and specific methodological advances. There is then a discussion of the strengths and limitations of the overall approach and study design. I address objective 4 of the thesis with a section on the policy implications arising from the studies, and the chapter concludes with a description of on-going research.

### Key findings with reference to objectives

#### **Objective 1: In people with CF, what is the relationship between SES and important longitudinal clinical outcomes, such as growth, and lung function?**

Study 1 addressed the first objective of this thesis, with an in-depth longitudinal study of the effect of SES on a range of clinical outcomes in the contemporary CF population in the UK. The key findings were:

*People with CF in the UK from more deprived areas, as measured using the indices of multiple deprivation for the UK constituent countries, have worse growth outcomes (weight, height and BMI) from around the time of diagnosis.*

Comparing the most deprived quintile to the least, the difference in weight and height SD-score is about one third, whereas for BMI the difference is about one sixth of a standard deviation score (z-score). These differences in anthropometric measures are statistically significant, and are also clinically significant. For instance, CF is a condition that has a profound effect on nutritional status in the first few years of life, and this was demonstrated in study 1 by comparing the mean weight SD score in the UK CF paediatric

population, to that of the UK reference population, a difference that equated to about a third of an SD score. This is the same as the deprivation effect, comparing extremes of deprivation quintile. Thus having CF, and living in the most deprived areas of the UK effectively doubles the nutritional disadvantage experienced by a person with CF.

***The deprivation gap in weight is greatest in the first few months of life, at the time of diagnosis, and there is a narrowing of the gap up to around age three.***

Study 1 quantified the age related changes in the deprivation gap in anthropometric outcomes, and showed that this narrows over the first three years of life. Extrapolating the modelled trajectory to around birth, the analysis suggests that the gap in weight SD score, comparing extremes of deprivation quintile, narrows from around -0.54 SD scores (95% CI -0.73 to -0.34) at birth to -0.28 (95% CI -0.38 to -0.18) at age three, and then the gap remains constant into adulthood. This analysis suggests a halving of the deprivation gap in the first three years of life. By contrast the analysis showed that the deprivation gaps for height (-0.31, 95% CI -0.40 to -0.21) and BMI (-0.13, 95% CI -0.22 to -0.04) did not change over time.

***People from the most deprived areas in the UK with CF have worse lung function when this can be routinely measured at around age five, but the deprivation gap does not increase after that age***

We currently consider %FEV<sub>1</sub> the most important clinical outcome in CF, because it is related to survival. Study 1 demonstrates that people from the most deprived areas have significantly worse lung function (-4.12 percentage points, 95% CI -5.01 to -3.19). Unlike some analyses from the US, the deprivation gap in %FEV<sub>1</sub> did not increase over the paediatric age range, but remained constant over time. The finding of a fixed %FEV<sub>1</sub> deficit in the most deprived children at age five has important policy implications, and suggests that deprivation has a detrimental effect on lung health in the early years of a child's life.

***The prevalence of chronic *P. aeruginosa* colonisation is higher in people from the most deprived areas of the UK, from an early age.***

Acquisition of *P. aeruginosa* can occur very early on in life in people with CF (Rosenfeld et al., 2012a), and becomes increasingly common over childhood. It is associated with worse outcomes, including survival, lung function, and nutritional status (Kosorok et al., 2001, Emerson et al., 2002, Konstan et al., 2007a, Taylor-Robinson et al., 2012a, Taylor-Robinson et al., 2013a). Study 1 demonstrated for the first time in a population wide cohort that *P. aeruginosa* acquisition appears to be related to social deprivation, such that people from more deprived areas are more likely to have *P. aeruginosa* across the age range (OR 1.89 in the <18 age group, 95% CI 1.34 to 2.66). Furthermore, this relationship persisted after adjusting for level of %FEV<sub>1</sub>. Acquisition of *P. aeruginosa* may be an important mediating factor in the relationship between social deprivation and poor outcomes in CF.

***In the adult population, growth and *P. aeruginosa* outcomes were worse in the most deprived adults, but no significant difference in lung function was detected***

Study 1 was the first to systematically document the effect of SES on an adult CF population. The lack of an association between SES and %FEV<sub>1</sub> in the adult population is perhaps surprising, given the tendency for social inequalities in health outcomes to increase over time, and the possible reasons for this are discussed in further detail in a subsequent section. These findings may relate to limitations of the dataset in the adult population, for instance due to left censoring (survival bias), and less power to detect an effect in the adult age range.

**Objective 2: In people with CF, what is the relationship between SES and health service use?**

Study 1 addressed the second objective of this thesis, with an in-depth longitudinal study of the effect of SES on use of therapies and health service related factors in the contemporary CF population in the UK. The findings were:

***There is no apparent difference in the age at diagnosis by SES for people with CF in the UK***

Study 1 assessed time to diagnosis by SES, and found no significant association. Furthermore, there was no significant social gradient in the proportion of individuals diagnosed by neonatal screening in the UK.

***People from more deprived areas of the UK with CF are MORE likely to use any IV therapy in the preceding year, after adjustment for disease severity***

Use of IV antibiotic therapy for the treatment of respiratory infections, and clinical exacerbations is a major treatment modality in CF. Study 1 demonstrated that people in the UK from more deprived areas are about twice as likely to receive IV antibiotics in a particular year, after adjustment for disease severity, on the basis of %FEV<sub>1</sub> and *P. aeruginosa* colonisation status (OR 2.52, 95%CI 1.92 to 3.17 in the <18 age group, and OR 1.89, 95%CI 1.51 to 2.38 in the adult age group, comparing most to least deprived quintile). This apparent positive discrimination persists across the age range, from infancy up to 40 years of age.

***People from more deprived areas of the UK with CF are MORE likely to have more days of IV therapy, conditional on having any IV therapy***

If children and adults with CF are recorded as having any IV therapy in a particular year, then people from more deprived areas in the UK, compared to the least, are more likely to have more treatment days, after adjusting for disease severity (15.9% more days in the paediatric age group, 95% CI 8.2 to 24, and 10.6% more days in the adult age group 95% CI 2.5 to 19.2).

***People from more deprived areas of the UK with CF are MORE likely to be treated with IV therapies in hospital***

Study 1 disaggregated the use of any IV antibiotics into use of therapies at home, and in hospital, and demonstrated that the positive discrimination in use of any IV therapy was almost entirely due to delivery of these therapies in

hospital, as opposed to home-based treatment. Overall, children from more affluent areas were more likely to be treated at home compared to their more disadvantaged counterparts.

***People from more deprived areas of the UK with CF are MORE likely to be treated with nutritional therapies***

Study 1 demonstrated that, as with the use of any IV therapy, people from the most disadvantaged areas of the UK are more likely to receive nutritional support (defined as receiving nutritional supplements orally, by nasogastric tube, gastrostomy tube, jejunal tube, or total parenteral nutrition), after adjustment for nutritional status, measured on the basis of BMI SD score (OR 1.78, 95% CI 1.42 to 2.2 in the <18 age group, and OR 2.38, 95% CI 1.69 to 3.36 in the adult age group, comparing the most to the least deprived quintile). This apparent positive discrimination in the delivery of nutritional support to the UK CF population was consistent across the age range studied, from infancy up to the age of 40.

***People from more deprived areas of the UK with CF are LESS likely to be treated with DNase and inhaled antibiotics***

Study 1 demonstrated inequality in the delivery of inhaled therapies to the UK CF population, after adjusting for measures of disease severity. Furthermore, this association becomes stronger in the adult age range. In children, there was no association between DNase use and SES, prior to adjustment for disease severity. However, after adjustment for disease severity, there was an apparent inequality, suggesting that DNase may not be delivered equitably in children. This association was clear in the adult population in both the unadjusted and adjusted association, with adults from more affluent areas being more likely to report using DNase or inhaled antibiotics in the preceding year.

*There was a ‘social gradient’ for most of the associations between deprivation status, and health and treatment use outcomes*

I have described and visualised the association between SES and outcomes in this thesis in the plots in terms of the gap between the least deprived and most deprived quintiles. However, for most of these associations (with the exception of BMI SD score, and DNase use in the paediatric age range) there was a graded dose response relationship, suggesting the existence of a social gradient.

### **Objective 3: What is the impact of SES on longitudinal employment status in people with CF?**

Study 2 addressed the third objective of this thesis, with a longitudinal study of the effect of SES and disease severity on employment chances in adults with CF in the UK. The main finding was:

*All other things being equal, people from more disadvantaged areas in the UK are less likely to be in employment, and furthermore, social deprivation modifies the effect of disease severity in CF: poor lung function, as measured by lower %FEV<sub>1</sub>, is more harmful to employment chances for people living in the most disadvantaged circumstances compared to the least.*

In summary, in relation to the main objectives of this thesis, the studies 1 and 2 have demonstrated that people from more disadvantaged areas in the UK have worse health and social outcomes. They are more likely to have poor growth, lung function, and to acquire *P. aeruginosa*, and are less likely to be in employment. In contrast to this, the use of major therapies in the UK CF population shows a so called ‘pro-poor’ bias, with people living in the most deprived areas of the UK around twice as likely to be treated with IV antibiotics and nutritional therapies, after adjusting for disease severity. However, there was evidence of inequalities in the use of home IV therapies, DNase, and inhaled antibiotics. The final objective of this thesis, assessing

the policy implications of the findings around the effects of SES on the CF population is considered in a subsequent section.

In order to determine the effect of SES on longitudinal outcomes in the UK and Denmark it was necessary to understand the overall changes in outcomes in the population, and the influence of other important covariates. This contribution to the knowledge base is described in more detail in the next section.

## **How has this thesis contributed to the literature?**

This section describes the substantive advances, and other key findings from the studies in this thesis, made in three areas: CF epidemiology; our understanding of health inequalities more general; and methodological advances. The findings are contextualised in relation to the current knowledge base.

### **Contribution to our understanding of CF epidemiology**

#### **Inequalities in outcomes in CF**

##### Establishing an association between SES and longitudinal health outcomes in CF in settings outside the US

Study 1 builds on the findings from the US of an association between SES, and adverse outcomes in CF (Schechter et al., 2001, O'Connor et al., 2003), by examining the full range of growth measures (weight, height and BMI) and lung function, and extends the US findings to examine longitudinal *P. aeruginosa* acquisition. Furthermore, study 1 demonstrates the presence of a social gradient in CF health outcomes, suggested in the O'Connor study for weight and %FEV<sub>1</sub> (O'Connor et al., 2003), and uses for the first time a validated measure of small-area deprivation.

The findings from study 1 constitute a novel contribution to the literature, since this study is the first to examine CF outcomes, other than mortality, in the UK, from a health equity perspective. Two studies in the UK have explored the effects of SES on risk of death (Britton, 1989, Barr HL, 2011). The Barr study – an update of the Britton analysis published since the start of this thesis – showed that the socio-economic divide in premature mortality in CF has persisted with no substantial narrowing for over four decades between 1959 and 2008. SES was measured on the basis of occupational social class, and in the most recent period the OR for survival above the median age at death was 1.89 (95% CI 1.20 to 2.97), comparing professional/managerial groups to routine/manual groups.

Other than these studies, there have been no population level studies in the UK, or other European countries, exploring the effect of SES on longitudinal outcomes in CF.

A concern with extrapolating the findings from the US to the UK and elsewhere is that any social differences in outcomes could be the result of the very different health care and welfare systems operating on either side of the Atlantic. However, the results presented in this thesis suggest that SES is indeed a ‘disease modifier’ in both the US and the UK (Schechter, 2013). Appealing to Bradford-Hill’s suggested second criteria for establishing causation, ‘consistency’ of findings in different places, and with different samples, strengthens the likelihood of an effect (Hill, 1965).

My on-going studies using the Danish registry, applying the methodology described in this thesis, will extend this knowledge base further, since it remains important to examine the extent of SES related inequalities in CF in other countries. This will open up the possibility of comparative studies, which may provide further insight into the precise mechanisms by which SES is exerting an effect on outcomes, particularly with regard to the influence of the macro-economic context.

#### Establishing a deeper understanding of inequalities in CF over the life-course

Study 1 is the first to use modern longitudinal data analysis techniques to quantify the age related changes in CF outcomes, stratified by SES, from infancy up to the age of 40 years. There are three important contributions to the literature discussed here: inequalities in outcomes from the outset; the finding of narrowing inequality in weight SD score in the first three years of life; and the lack of increasing inequality over time in the UK.

The first important insight is the presence of inequalities that are evident from the time of diagnosis in the UK. Michael Schechter picks up on this in his editorial linked to the published version of study 1 (Schechter, 2013, Taylor-Robinson et al., 2013a):

*“[David Taylor-Robinson’s] analysis substantiates previous reports of decreased weight, height, and body-mass index, and worse lung function in patients with cystic fibrosis living in the most disadvantaged areas, but also elucidates several additional details of interest. First is that deprivation seems to have its greatest effect on somatic growth and lung function in early childhood; disparities are maintained but do not seem to increase as children age. Infancy and early childhood are a period of vulnerability to adverse exposures associated with poverty, which suggests that interventions to mitigate such effects might be most effective and efficient if focussed on the early years of childhood.”*

The presence of significant differences in outcomes by SES early in the life-course were suggested in the US studies, and have been confirmed, and quantified in the analysis in study 1 of this thesis. In a cross-sectional analysis, Michael Schechter’s study reported a greater chance of being under the 5<sup>th</sup> centile for weight or height in children using Medicaid in the US, and adjusted for age, but did not model age related changes in these outcomes. Lung function in Schechter’s study was modelled as a linear function of age, but this was limited to the cross-sectional relationship (Schechter et al., 2001). O’Connor et al demonstrate a social gradient in weight percentile at around diagnosis in the US, and in %FEV<sub>1</sub> from age five onwards, but do not develop a parametric model for the age related changes in these relationships (O’Connor et al., 2003).

The finding of a social gradient in growth outcomes, evident from around the time of diagnosis, points to important effects of SES in utero, and/or in the initial period prior to diagnosis. Both are plausible, but a limitation of studies thus far has been a lack of data on birth-weight, which would complete the picture. As discussed in chapter 3, the association between low SES and low birth-weight, and length is well established in the general population (Spencer et al., 1999, Marmot et al., 2010b, Howe et al., 2011, Howe et al., 2012). The causes for this are likely to be multi-factorial and relate to maternal health, including stress, diet, drug, alcohol and tobacco use during pregnancy (Marmot et al., 2010b). The question remains as to how having CF modifies this relationship. We can speculate that there may be an interaction, and that exposure to tobacco smoke *in utero* may be an important

influence. The importance of ETS exposure is discussed in more detail later on in this section.

Direct comparison of the SES effect on growth in the CF population, and the general population is challenging, due to the lack of studies using comparable SES measures and birth cohorts, and due to the effect of the obesity epidemic in the general population, whereby a relationship between overweight and low SES emerges from around age four onwards (Howe et al., 2011, Howe et al., 2012). This is in contrast to CF, where children are consistently underweight compared to the UK reference population (Taylor-Robinson et al., 2013a).

Further data on SES gradients in birth-weight in CF would clarify the extent to which the early growth differentials demonstrated are simply a reflection of the broader SES effects on birth-weight in the general population. Infants with CF not diagnosed by screening are generally underweight at the time of diagnosis in the US and Canada (Lai et al., 1999), and in the UK, as demonstrated in study 1. The UK registry does not capture differential weight loss by SES in the period prior to diagnosis, and this may be particularly important in the first few weeks of life, compounding any influence of SES on birth-weight. Data on these very early growth trajectories in CF, coupled with birth-weight data is emerging from recent studies in the US of babies diagnosed by new-born screening (Lai et al., 2009, Jadin et al., 2011), but these are limited to small samples, and do not examine SES effects. It is likely that studies like these, focussing on babies diagnosed at birth, hold the key to unlocking this part of the puzzle, and the rollout of universal newborn screening in the UK offers great possibilities for examining this further.

I hypothesise that newborn screening in the UK may reduce inequalities in CF outcomes. This was explored in the supplementary analysis of the effect of SES on weight trajectories, in the screened and unscreened population, presented in Chapter 3 (Figure 36). The point estimates suggested a narrowing of the SES related weight gradient in the screened population of the UK, though these were not statistically significant. We can speculate that early diagnosis might reduce any early differential weight loss by SES, especially when coupled with more intensive delivery of nutritional therapies to children from more disadvantaged areas demonstrated in study 1. As the number of screened children in the UK increases, the power to test

this hypothesis further will improve.

This hypothesis is further supported by the second novel contribution to the literature, relating to the narrowing of inequality in weight SD score in the first three years of life. This suggests that diagnosis and treatment for CF in the NHS may be having a pro-poor, or ‘levelling-up’ effect, and further we can speculate that earlier diagnosis by screening would increase this effect. The third important finding, closely related to the decrease in inequality in weight SD score, is the lack of widening inequalities over time across all clinical outcomes studied in the UK. The O’Connor study in the US suggested that this may be the case, as their analysis also did not suggest an age related increase in weight percentile and %FEV<sub>1</sub> inequalities by zip-code linked income (O’Connor et al., 2003), whereas Michael Schechter’s cross-sectional analysis suggested an increase in the age related deprivation gap in %FEV<sub>1</sub> on the basis of Medicaid status (Schechter et al., 2001).

The findings of decreasing inequality in early life for weight, and stable inequality subsequently for other outcomes are important. The life-course perspective suggests that the accumulation of risk with age tends to lead to increasing inequalities over time. For instance, the UK Marmot review states (Marmot et al., 2010b):

*“Central to the Review is a life course perspective. Disadvantage starts before birth and accumulates throughout life.”*

Furthermore, the analysis in study 1 did not demonstrate an effect of SES on %FEV<sub>1</sub> in the adult population, and though the effect of SES on BMI SD score was of the same magnitude in adults and children, it just failed to reach significance (−0.12 SD scores, 95% CI −0.25 to 0.01, comparing most to least deprived quintile).

There are a number of possible explanations for these findings. It may be a real phenomenon, and we can speculate that the pro-poor delivery of key therapies in CF (IV and nutritional therapies) is having a levelling-up effect, and that in the absence of the more intensive delivery of treatments to disadvantaged children, the counter-factual situation would be for inequalities in outcomes to increase with age. If this is the case, this is important, as it provides evidence to suggest that health services can have a levelling-up influence on chronic diseases. The findings may, on the other

hand, be artefactual, possibly as a result of factors such as left-censoring in the dataset, selective dropout due to death, and reduced power to detect effects in the adult age range due to the smaller numbers of individuals, and the relatively short longitudinal data traces in the UK study. As argued in the methods section, the UK dataset, compared to the Danish data, is better suited to demonstrate cross-sectional differences between patients, rather than change over time. These issues are considered in more detail in the limitations section. This study is the first to explore SES effects in adults with CF at a population level, and our understanding of the effect of SES across the lifespan on clinical outcomes in CF would benefit from further studies in this age group (Schechter, 2013).

#### Establishing a better understanding of the role of health services in the pathway to health inequalities in CF

A key question is the role of health services in perpetuating or remediating against the generation of any health inequalities demonstrated in CF. Margaret Whitehead defines equity in healthcare as:

- equal access to available care for equal need;
- equal utilization for equal need;
- equal quality of care for all (Whitehead, 1990).

Study 1 thus explored aspects of CF care in the UK from an equity perspective, and overall the picture is mixed. There is no difference by deprivation status in age at diagnosis, or in the degree to which patients with CF are captured in the registry. This suggests that children with this chronic disease are entering the care system, and accessing the annual review follow-up process in an equitable manner. With regard to the major treatments for CF, study 1 provides evidence of more intensive treatment being delivered to both children and adults with CF, for two major pillars of CF care – treatment with IV therapies, and nutritional therapies – after adjusting for measures of disease severity. These findings provide evidence of a so-called ‘pro-poor’ bias (Ravallion, 2001). Study 1 also demonstrates that children and adults from more disadvantaged areas are more likely to be treated in hospital with IV therapies, again after adjusting for disease severity. Aside from consideration of residual confounding by severity, one explanation for these findings is that clinicians are actively taking steps to address, and overcome the perceived excess disadvantage

faced by people living in more deprived areas. This explanation rings true with many of the paediatricians with whom I have discussed these results. Some have suggested that they take a more paternalistic approach to children with CF who are perceived to be living in more disorganised home circumstances, where there may be additional barriers to ensuring that they receive the care and treatments needed. In these situations, some clinicians have suggested they prefer to admit children to hospital, where a range of treatments can be delivered with certainty.

This line of reasoning is corroborated in a recent case-report regarding a child with CF under the care of the Brompton hospital (Gupta et al., 2009). The article describes the dramatic improvement in the clinical condition of an eight-year old girl, following a change in her challenging home circumstances. The authors describe how she was regularly admitted to hospital for treatment optimisation and intensive physiotherapy, but that within weeks of discharge from hospital the benefits of her treatment would 'wear off'. Her home circumstance changed following the imprisonment of her mother and brother, when she went to live with an older sister, and this coincided with a dramatic improvement in lung function, nutritional status, and reduced treatment requirements. The authors of the case report suggest that it may be appropriate to offer children living in more adverse conditions more intense standard treatment including additional supervised antibiotics, and physiotherapy, and further suggest that the efficacy of such approaches needs to be assessed (Gupta et al., 2009). The findings from study 1 suggest that such practices may already be widespread in the UK, and we can further speculate that this may be having a positive impact on the clinical condition of children from the most disadvantaged areas, as manifested by the reduction in weight inequality over the first few years of life, and the relatively stable inequalities gap demonstrated in study 1, as opposed to a finding of increasing inequality.

By contrast, the finding of inequality in the delivery of important inhaled therapies such as DNase, and inhaled antibiotics is of immediate concern, and requires further investigation. It is unclear if these differences are intended or not. These therapies are time-consuming to administer, and need to be taken consistently to deliver clinical benefits, with a recent systematic review demonstrating improvements in lung function from long term DNase therapy (Jones and Wallis, 2010). These differences may relate to conscious decisions made by clinicians relating to the

balance between the likelihood of benefit, and the burden of therapy in CF. Clinicians may be intentionally less likely to prescribe these long-term inhaled therapies in an attempt to reduce the burden of care if they feel that patients are less likely to benefit, because of concerns over adherence, or because patients actually report not taking them. The high burden of care is well documented for people with CF (Ziaian et al., 2006, Sawicki et al., 2009), and the addition of numerous aerosol treatments to physiotherapy routines means a treatment burden in excess of two hours per day for many individuals with CF (Geller and Madge, 2011). Furthermore, a recent study by Quittner et al (Quittner et al., 2010) has suggested that lower SES in the US, as measured by Medicaid status, is associated with a perceived higher burden of care in families of children with CF, after adjusting for disease severity. Added to this is the high non-compliance rate to treatments in CF, and the fact that this is likely to be related to SES (Schechter, 2011). Studies of non-adherence in CF have suggested rates ranging from 20% to 70% in the UK, US and Spain depending on the type of treatment and the method of measurement (Arias Llorente et al., 2008, Weiner et al., 2008). In addition compliance has been found to worsen with age and disease severity (Arias Llorente et al., 2008), which could explain the greater inequality in the adult population.

On the other hand, these differences may be unintended, and reflect the marked variations in prescribed CF care that have been described in the US (Konstan et al., 1999b, Konstan et al., 1999a, Schechter and Margolis, 2005) on a centre basis, that are also likely to occur in the UK. Study 1 provides some evidence of this, whereby adjusting for care centre effects reduced the SES effect on risk of any IV therapy use somewhat. Patients and families from more disadvantaged areas may be more susceptible to these unintended inconsistencies in the delivery of care because they are less able to 'navigate' the care system, due to reduced self-advocacy or self-management skills (Schechter, 2013).

Schechter et al have conducted two further studies aimed at exploring the differences in use of CF related treatments in the US (Schechter et al., 2009, Schechter et al., 2011). In the paediatric population of the ESCF of around 10,000 children, these studies confirmed the tendency for children on Medicaid to be more likely to access treatments, whereas the picture was mixed when using other SES indicators such as area-based income. This is a further indication that access to free health care is

important in the ability to uptake appropriate treatments. For instance there was no difference in IV therapy use in children <12 by area-based income, but in young people aged 13-18 years, those living in more affluent areas were more likely to be treated (13.8% in the lowest income category compared to 19.2% in the highest) (Schechter et al., 2011).

In contrast, another study from the US focussed on the association between SES and the chances of being accepted for lung transplant in 2167 adult patients with CF evaluated for transplant in the US, between 2001 to 2009 using data from the CF Foundation Patient registry (Quon et al., 2012). The authors found that people using Medicaid (low SES) were less likely to be accepted for lung transplant compared to people without Medicaid, and this association remained after adjustment for demographic factors, measures of disease severity, and contraindications to transplant. A similar association was found when using alternative methods of SES, including educational attainment and area-based household income.

Interpretation of these studies that use Medicaid status in the US as a measure of SES, is more difficult when the outcome is health service use, as opposed to health status, since Medicaid use in the US is both an indicator of low SES, and an indicator of free access to health care. However, the general finding from the US studies is that children using Medicaid are also more likely to access treatment, with the exception of lung transplantation, after adjusting for need (Schechter et al., 2001, Schechter et al., 2009, Schechter et al., 2011).

In summary, study 1 suggests that the NHS in the UK may be having an important influence on inequalities in outcomes in CF, in ways that differ across the life-course, depending on the treatment modality. We can contrast these findings with those from the US, where studies have not demonstrated major differences in uptake for common CF treatments, with the exception of transplantation, and have tended to play down the role of health services in the generation of health inequalities.

### Demonstrating differential social consequences in CF

The findings from study 2 are novel. No previous studies have explored the interplay between SES, disease severity, and employment chances in people with CF, in the UK or elsewhere. People with CF are living longer so it is important to understand factors that influence adult employment chances. Previous studies have shown that people with CF from socio-economically disadvantaged backgrounds have worse outcomes (Schechter et al., 2001, O'Connor et al., 2003, Taylor-Robinson et al., 2013a), and study 2 builds on these findings to demonstrate how SES, disease severity, and time in hospital influence employment chances, in order to better understand how health and social inequalities in CF might be perpetuated.

Around 50% of adults with CF were in employment. Lower social deprivation, male sex, higher %FEV<sub>1</sub> and BMI, and less time in hospital were associated with improved employment chances. Crucially, study 2 demonstrates for the first time that deprivation modifies the effect of disease severity in CF: poor lung function is more harmful to employment chances for people living in the most disadvantaged circumstances compared to the least. Low employment in people with CF is a serious concern. Being out of work increases the risk of poverty and social exclusion, and is likely to further damage the health of the most disadvantaged people with CF. In study 2 I have demonstrated *differential social consequences of illness* in the context of CF, by which people with the double burden of chronic illness and low SES are more likely to be excluded from the labour market. We can speculate that this may be an important pathway for the amplification of health inequalities in CF, whereby disadvantage builds on disadvantage. It is of particular concern that the most disadvantaged women have the poorest employment chances, since female sex is also an important risk factor for poor outcomes in CF (Barr HL, 2011, Taylor-Robinson et al., 2013a). These findings suggest that policies should be considered to address the extra burden of adverse consequences of CF faced by patients living in disadvantaged circumstances.

## Applying the Diderichsen model to CF

The studies in this thesis were informed by the Diderichsen model of pathways to health inequalities (Diderichsen et al., 2001). Taking each mechanism in the pathway in turn, we first consider differential exposure. The key feature of CF as a model to study health inequalities is that the risk of having two defective CF genes is not significantly socially graded, and therefore differential exposure to this risk factor does not explain inequalities in CF outcomes. However, I have demonstrated differential exposure to *P. aeruginosa* acquisition by SES, which is a key risk factor for disease progression in CF (Taylor-Robinson et al., 2013a).

Study 1 provides an important example of differential vulnerability. This pathway hypothesises that even if a risk factor is evenly distributed by SEP, it can result in differential effects due to the synergistic action of other exposures. In CF, we are seeing differential consequences in terms of clinical outcomes, stemming from the same exposure to the underlying genetic risk factor for CF. This suggests that there is differential vulnerability in terms of having two CFTR mutations, on the basis of SES, which is likely to be due to the interaction of other exposures that are socially graded (e.g. second-hand smoke exposure, nutrition, stress, environmental exposures) that I have been unable to quantify using the data available in the CF registry. I will discuss these in further detail in the next section.

As a result of the twin pathways of differential exposure, and differential vulnerability, I have clearly demonstrated differential health consequences in CF in terms of lung function, and growth (Taylor-Robinson et al., 2013a), and the studies by Britton and Barr suggest that these factors are likely to lead on to important differences in survival in the UK (Britton, 1989, Barr HL, 2011). Finally to close the loop in the Diderichsen model, study 3 demonstrates differential social consequences, whereby poor lung function is more damaging to employment chances in more disadvantaged areas.

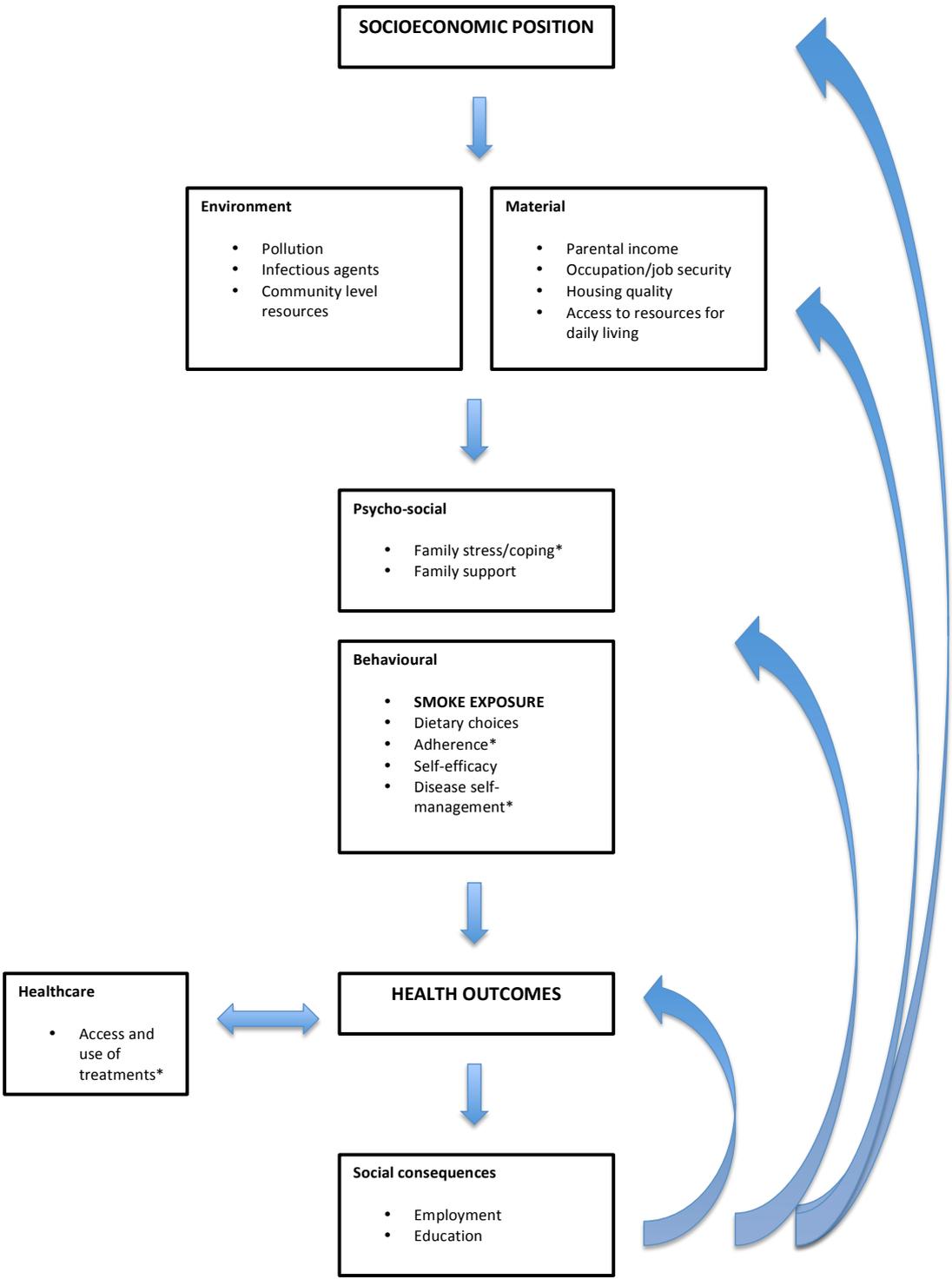
## **Understanding inequalities in CF**

Having discussed the contribution of the studies in this thesis to our understanding of inequalities in CF outcomes, this section concludes with an overview of our current understanding as to the mechanisms and pathways that are likely to be important in causing the differences in outcomes that I have documented. Whilst I have been able to demonstrate the presence of differential exposures, vulnerability, outcomes, and consequences as discussed in the section above, there remain key exposures that are likely to be important that are not captured in CF registries. We can speculate about their likely relationship with SES, and adverse outcomes in CF, with reference to the broader literature.

Figure 78 is an attempt to illustrate a range of exposures that are likely to be important in influencing CF outcomes, and that are likely to be differentially patterned by SES. The vertical ordering of the exposures aims to represent concentric layering of influences as per the Whitehead rainbow (see Chapter 2), whereby broader ‘upstream’ socio-economic factors influence the distribution of the main determinants of health, which further influence ‘downstream’ behaviours. Thus parental occupation and income will influence access to a good diet, and the dietary choices that are possible.

The exposures are also distributed over time, and as I have argued, exposures in the early years may be particularly important. Furthermore, these effects may interact with one another over time, in ways that are difficult to entangle. For instance, I speculate that the intensive delivery of certain therapies to disadvantaged children may act to mitigate some of the adverse effects relating to the usual picture of accumulating health damaging exposures over time. The following sections consider the potentially important risk factors that have not been discussed in detail previously in this thesis, and that are not captured in the registry data, starting with tobacco smoke exposure, since this is likely to be particularly important. Health care has been discussed earlier in this chapter, and is not considered in any further detail in this section.

**Figure 78: Logic model based on the Diderichsen model outlining exposures for poor outcomes in CF**



### **The importance of tobacco smoke exposure**

We can speculate that tobacco smoke exposure *in utero*, and in the early years may be an important mediator of the relationship between SES and adverse outcomes in CF, since there are striking and persistent differences in smoking prevalence by SES in the general population (Marmot et al., 2010b), and ETS exposure is associated with poorer growth and lung function in CF (Rubin, 1990, Kovesi et al., 1993, Schechter, 2004).

In 1990, Bruce Rubin first demonstrated a dose-response relationship between ETS exposure and poor lung function and growth in children with CF attending a summer camp in the US (Rubin, 1990). A US study of data collected since 2000 looked in detail at 800 participants in the US Cystic Fibrosis Twin and Sibling Study with data collected on ETS exposure (Collaco et al., 2008). In this study 23% of patients were exposed to ETS in the home and 17% were exposed to tobacco *in utero*. Furthermore exposure in the home was associated with significantly worse lung function, but not BMI, and *in utero* exposure was significantly associated with lower birth weight by 150 g. The authors suggest that the detrimental effect of secondhand smoke exposure is equal to that expected for an additional 7.3 years of lung function decline. The study also suggested gene-environment interactions between ETS exposure non delta F508 CFTR mutations, and the CF modifier gene (TGF), with individuals carrying these genes being more at risk of the adverse effects of ETS exposure. Unusually, the authors did not find a relationship between SES (measured on the basis of parental education, and ZIP code linked income) and exposure to ETS, which suggests that the study population may not be representative of the general CF population. Furthermore they did not find a consistent independent association between SES and clinical outcomes, once ETS exposure was added to their regression model – there was an association with paternal education, but not for maternal education level or income. On this basis the authors suggest that SES and ETS should be considered as confounding one another, but I suggest that it is more appropriate to consider ETS exposure as a mediating factor, as in the social epidemiology literature (Taylor-Robinson et al., 2011) (Morgen et al., 2008).

In terms of data from the UK, a study of smoking in CF families in one UK centre showed a high prevalence, that did not change substantially over the period of

measurement: In 1993, 26/56 (46%) households contained at least one smoker (smoking households) compared with 23/52 (44%) in 1998 (Smyth et al., 2001).

### **Environmental exposures**

Several studies have suggested that outdoor air pollution may have an important impact on the progression of CF lung disease (Goeminne et al., 2012, Farhat et al., 2013, Jassal et al., 2013). A US registry study linked annual average exposures to particulate air pollution to increased risk of pulmonary exacerbations and a decline in lung function, suggesting a role of environmental exposures on prognosis in CF (Goss et al., 2004). Low SES has been linked to increased risk of exposure to pollution in England (Wheeler and Ben-Shlomo, 2005). Furthermore in the UK general population, low social class and poor air quality were independently associated with decreased lung function, and there was evidence of differential vulnerability, whereby the adverse effects of air pollution seem to be greater in men in lower social classes. It is reasonable to speculate that these effects might be amplified by the presence of a chronic lung disease like CF. Further studies using the UK registry could usefully explore linkage to datasets capturing information on air quality to investigate these associations further.

Study 1 in this thesis has provided evidence of differential exposure to *P. aeruginosa*, for the first time in a population wide cohort. In Schechter and colleagues study (Schechter et al., 2001), Medicaid patients were more likely than non-Medicaid patients to have *P. aeruginosa*, but when adjusted for %FEV<sub>1</sub> there was no statistical difference, and a more recent US cohort study did not demonstrate an association (Rosenfeld et al., 2012a). *B. cepacia* is another important organism in CF, shown to be associated with increased morbidity and mortality (Ledson et al., 2002). In study 1, *B. cepacia* was associated with increased rate of decline in lung function, but the small numbers of people affected precluded analysis by SES as a main outcome (156/4445 in the %FEV<sub>1</sub> analysis <18 age group). It remains to be seen if low SES is a risk factor for *B. cepacia*. Like *P. aeruginosa*, however, exposure to other patients with the organism is a risk factor for acquisition, and thus increased time spent in hospital (associated with low SES in study 1) may also be a risk factor (Ledson et al., 2002). In a similar manner, exposure to viral respiratory infections may be associated with adverse outcomes in CF (Hiatt et al., 1999, Asner

et al., 2012), and a number of studies have suggested a relationship between low SES and more severe consequences of early respiratory viral infection (O'Donnell et al., 2010, Leem et al., 2011).

'Community level resources' is a catch-all phrase, but would include factors such as access to green space, and safe play areas in the local community, which may influence the relationship between SES and CF outcomes. For instance, access to green space has been related to levels of exercise and wellbeing in the general population. Furthermore there is a clear social gradient in exercise levels by SES (Geddes et al., 2011). In addition, we know that exercise has clear benefits in CF (Schneiderman-Walker et al., 2000).

### **Material resources**

Material factors, directly related to level of income, are likely to have important effects on children and adults with CF at many levels, which are complex and difficult to capture completely. In the general population, for instance, insufficient income is associated with worse outcomes across virtually all domains, including long-term health and life expectancy (Marmot et al., 2010b). The Marmot review accordingly recommends the development of the concept of a minimum income for healthy living (Morris et al., 2000, Morris et al., 2007), which would ensure that all would receive an appropriate income for their stage in the life course, and would reduce overall levels of poverty as well as child poverty (Marmot et al., 2010b). With regard to CF, the influence of income on housing quality and access to a healthy diet are illustrative examples. Low income is linked to poor housing quality, with the presence of damp and mould, which in turn have been linked to adverse respiratory outcomes in diseases such as asthma (Krieger and Higgins, 2002). Nutritional status is an important influence on lung function and survival in CF, and study 1 has demonstrated an association between SES and nutritional status. Furthermore there are marked social differences in the quality of dietary intake in children in the UK by SES, with significantly lower rates of breast-feeding and access to a healthy diet in more disadvantaged areas (Marmot et al., 2010b).

As well as material factors influencing health, it is clearly possible for a CF diagnosis to have a detrimental impact on family income. Furthermore, study 2 has

demonstrated differential social consequences for adults by SES. These pathways are also likely to be important for children, whereby a diagnosis of CF in a family imposes additional financial burdens. The literature in this area is limited, and economic analyses have tended to focus on the cost of CF to health care systems (Krauth et al., 2003), or the burden of out-of-pocket expenses to families in the context of the US health care system (Lieu et al., 1999). However, there is a body of literature relating to other severe illnesses in childhood, demonstrating financial stress, employment loss, and overall a negative economic impact on families with a critically ill child (Jacobs and McDermott, 1989, Montgomery et al., 2002, Winthrop et al., 2005, Shudy et al., 2006). In CF, three older studies touch on this issue. A diagnosis of CF in the family has been shown to increase household expenses in Norway (Folleras et al., 1988). Another study from the US notes the significant hidden costs that follow a diagnosis of CF in the family, including expenses associated with clinic visits, loss of pay to either or both wage earners, car expenses, parkway tolls and fees, and cost of childcare for siblings of the patient (McCollum, 1971). Furthermore, a small study in Norway also demonstrated striking differences between mothers of children with CF and controls with regard to their possibility of having a career. They found that while the education and professional career of the fathers of CF children were generally not influenced by having a chronically ill child to care for, mothers felt compelled to give up or adapt their education or professional career (Michalsen et al., 1988).

### **Psycho-social factors**

There can be no doubt regarding the stress that a diagnosis of CF places on families, and it is highly plausible that these effects might be amplified by social disadvantage. Stress pathways have been shown to be important in the generation of health inequalities (Lynch et al., 2000, Marmot and Wilkinson, 2001), and in worse health outcomes in CF (Yohannes et al., 2012). However, of potentially greater importance for children with CF, is the parental experience of stress, which may lead to anxiety, depression (Cruz et al., 2009, Besier et al., 2011), and impaired personal and family function (Patterson et al., 1993, Schechter, 2011). A recent study of the ESCF population explored health related quality of life in children and families with CF, by SES, and showed differential effects (Quittner et al., 2010). This study concluded

that children with CF and their parents from more disadvantaged backgrounds, measured by Medicaid status, appear to have worse quality of life scores across all domains of the Health related quality of life (HRQOL data), using the Cystic Fibrosis Questionnaire-Revised (CFQ-R), even after adjustment for disease severity. Furthermore, a recent review of the effects of paediatric illness on families indicated that pediatric critical illness is stressful for the entire family. The review, which included 115 studies, suggested negative effects on parents, siblings, and marital cohesion. The needs of family members in terms of rest, nutrition and communication were identified as being compromised in many studies. Furthermore, a permanent detrimental impact on siblings and marital relationships was described in some studies (Shudy et al., 2006).

### **Behavioral factors**

At the outset, I should make it clear that I recognise the cultural explanation of behaviour and how it influences inequalities, as described in the Black report. Here particular behavioural patterns are seen as being “embedded within the social structure” (Black et al., 1980), and are influenced by the material and psycho-social pathways described above. Thus dietary choices are restricted in disadvantaged areas, as discussed above. Furthermore stress leads to harmful health behaviours such as smoking (Kawachi et al., 2002). It is wholly understandable that the stress associated with dealing with a diagnosis of CF may increase the desire to have a cigarette in a parent who is a smoker, and the negative health consequences of this have been discussed above. Increased family stress may also directly influence factors such as adherence to medications, and the capacity for parents to adequately manage a child with CF. These issues were discussed in the previous section, where it was noted that the perceived burden of care is greater for more disadvantaged CF families (Quittner et al., 2010), and non-compliance rates are also likely to be higher (Schechter, 2011).

In summary there is evidence to support a number of plausible pathways that may help explain the differential outcomes demonstrated in study 1 (Taylor-Robinson et al., 2013a). Taken together, this literature should heighten awareness of the greater health and social risks faced by patients who are economically disadvantaged, and appropriate policy responses are discussed in a subsequent section.

## Understanding longitudinal growth in CF

Study 1 has described for the first time in a UK wide cohort the longitudinal growth trajectories for children and adults with CF. Furthermore, the study demonstrates the influence of important covariates on growth, such as sex, genotype, and screening status, as well as time-varying factors such as acquisition of *P. aeruginosa*.

I describe the changes in weight and BMI in the first few years of life, whereby the population is markedly underweight at the time of diagnosis, and there is a period of improving nutritional status over the first three years of life, presumably secondary to diagnosis, and consequent attention to nutritional status and pancreatic enzyme replacement. There were important covariate effects, with higher weight SD score associated with male sex, screened patients, heterozygotes for delta F508 and white patients in the <18 age group (Figure 34), and with female sex, heterozygote status and white patients in the >18 age group. Higher BMI was associated with male sex in the paediatric age range, and had a steeper rate of decline in delta F508 homozygotes after the age of three, and was consequently associated with lower BMI SD score in adults.

To date, this study represents the most comprehensive analysis of the dynamics of growth over time in a contemporary CF population, using modern longitudinal analysis techniques (Salvatore et al., 2012). Two studies have previously described longitudinal growth in the US and Canadian populations (Lai et al., 1999, Zemel et al., 2000). Zemel studied growth in 1000 children in the US with PI, using mixed-effects models, but with a limited range of covariates, and showed that important sex-related differences in growth occur before puberty, with males faring better. The authors also correlate better nutritional status with improved lung function after age six (Zemel et al., 2000). Lai et al compared the entire US and Canadian CF populations, using a GEE approach, and they demonstrate early malnutrition around the time of diagnosis, that improves over the first few years of life in both populations, with slightly better outcomes in the Canadian cohort, where mean height and weight were 4 to 5 percentiles higher than those in the United States (Lai et al., 1999). Lai et al also demonstrate a male and heterozygote (delta F508) advantage with regard to weight, but this is on the basis of a stratified analysis, rather than the multivariable approach in study 1 in this thesis.

In this thesis I have developed a parametric longitudinal model describing growth trajectories in the UK, which builds on these previous findings, and describes the effect of SES on early weight trajectories. Current studies of growth are particularly focussed on the first few months of life. Lai's group, for instance, have undertaken a number of more recent studies of growth, using smaller populations identified through the Wisconsin neonatal screening programme, with a focus on the longitudinal dynamics of early nutritional trajectories, and how these relate to subsequent outcomes (Shoff et al., 2006, Lai et al., 2009, Jadin et al., 2011). As in study 1 presented here, diagnosis through screening leads to an early nutritional advantage, and these studies are focussed on explaining why some patients with CF respond to treatment initiation and succeed in recovering from malnutrition and growth faltering/failure experienced before diagnosis, whereas others fail to do so. For instance, in the Wisconsin studies early attainment of birth weight status after CF diagnosis is associated with significantly better pulmonary outcomes at 6 years of age (Lai et al., 2009), and exclusive breastfeeding in the first few months of life was associated with adequate growth and protected against *P. aeruginosa* infections during the first two years of life in CF infants who had PI (Jadin et al., 2011). Furthermore a recent study of children diagnosed by newborn screening in Australia demonstrated a longitudinal association between markers of pulmonary inflammation and worse nutritional status, though the direction of this relationship requires further investigation (Ranganathan et al., 2011).

Height has been independently related to survival in people with CF, and it has been suggested that this may be mediated through greater lung capacity in taller people (Beker et al., 2001). Furthermore, Fogarty et al have suggested that measures of body habitus may partially explain the male survival advantage in CF (Fogarty et al., 2011, Fogarty et al., 2012). The findings from the analysis presented in this thesis suggest that SES has an important effect on height from the outset, and that this difference tracks through to adulthood, without the age related dynamics seen for weight and BMI. Further longitudinal analyses are required to understand how birth length, and the factors that influence it, may influence subsequent clinical outcomes and survival in CF.

In summary, malnutrition and poor growth are major concerns in children with CF, and optimizing nutritional status appears to be important for subsequent lung health.

Study 1 in this thesis has provided a longitudinal description of growth in the UK CF population. Future studies in the UK should develop our understanding of the changes in nutritional status in the early years, particularly utilising data from the growing number of screened children in the UK, taking into account SES as an important early influence.

### **Understanding changes in lung function in CF**

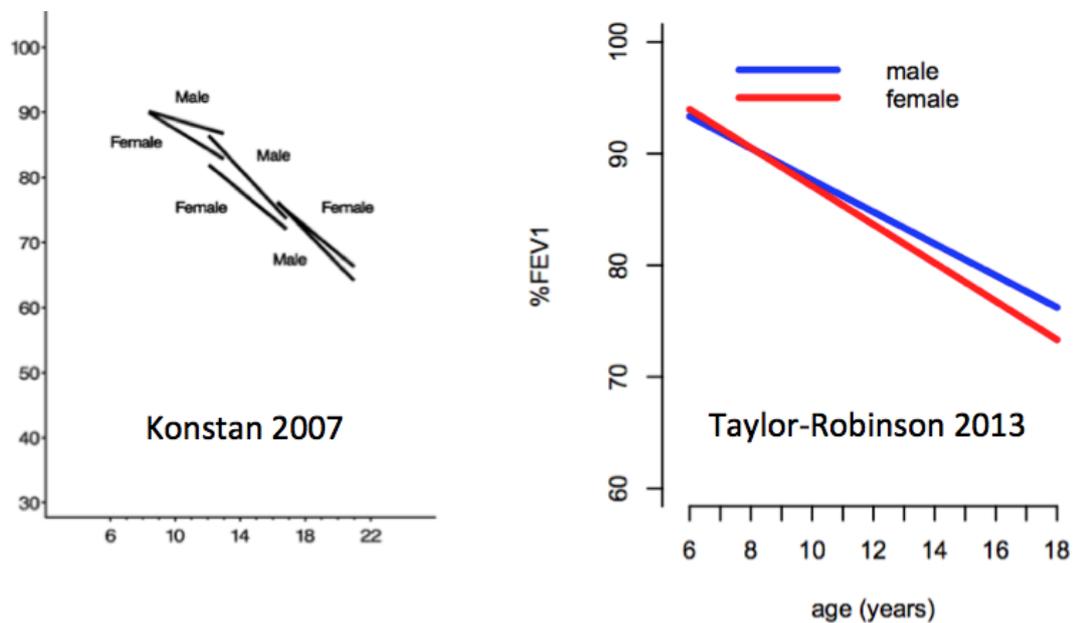
Study 1 and study 3 have added to the extensive body of literature on lung function decline in people with CF, introduced in Chapter 1 of this thesis, by describing factors influencing longitudinal lung function decline in both the UK and Danish population.

Turning to the UK population first, study 1 modelled the effect of a range of covariates on lung function decline in the paediatric and adult population. In the paediatric population, a linear model for the population average change in lung function provided a suitable fit to the data, and the effect of covariates was assessed using a random intercept and slope approach. In the paediatric population, higher %FEV<sub>1</sub> was associated with higher SES, male sex, screened patients, heterozygote delta F508 status, white patients, absence of CFRD and *P. aeruginosa* colonisation, and higher BMI. In the adult population, the population average was modelled as a split line, with a greater decline in lung function up to the age of 25, and there were similar significant associations between sex, and genotype, as in the paediatric analysis.

Konstan et al have undertaken the largest studies to date of %FEV<sub>1</sub> in both paediatric and adult cohorts (Konstan et al., 2007a, Konstan et al., 2012, Vandenbranden et al., 2012). In their first study of 4923 children in the US ESCF, using a mixed-effects regression approach, Konstan et al showed that higher baseline %FEV<sub>1</sub>, *P. aeruginosa* colonisation, female sex, and poor nutritional status were amongst the factors associated with a greater decline in lung function (Konstan et al., 2007a). In this study, the analysis was stratified into three different age groups (6 to 8 years, 9 to 12 years, and 13 to 17 years), and whilst the covariate effects were broadly similar over the age range, the population level rate of decline in %FEV<sub>1</sub> was shallower in the 6 to 8 age group, suggesting a population level acceleration in lung function

decline after the age of eight years. The study presented in this thesis identifies similar risk factors to those in the Konstan study, and although the models are not directly comparable, the findings are broadly consistent. Figure 79 below, for example, compares the effect of female sex on lung function decline in the US and UK paediatric population, with consistently worse lung function in females in both populations before the age of 18 years.

**Figure 79: Effect of sex on lung function decline in the US and the UK**



Vandenbranden et al published a study in 2012 focussing on the same ESCF population, exploring influences on lung function decline during the transition from late adolescence to early adulthood, since they hypothesized that there may be identifiable risk factors in adolescence associated with accelerated lung function decline during early adulthood that are amenable to intervention (Vandenbranden et al., 2012). Risk factors for substantial decline in early adulthood included slower rate of %FEV<sub>1</sub> decline in adolescence, poor nutrition, *P. aeruginosa* and male sex. This corroborates the findings in our study, and it is notable that male sex was associated with an increased risk of substantial decline in the early adult period, as in Figure 48 in my analysis of adults in the UK. There is the potential in the UK data to study these dynamics in more depth, with a greater focus in age related changes, but this was beyond the scope and purpose of the studies presented here.

In 2012 Konstan et al have published a study of the adult ECFS population (4161 adults). Together with the data presented in study 1 (4026 adults) these represent the largest studies of lung function decline in adult CF populations, since previous studies of risk factors for lung function decline have not focussed specifically on adults with CF, or did not evaluate adults separately by age groups, possibly due to small numbers of patients available for analysis (Konstan et al., 2012). Although study 1 in this thesis explored different risk factors, there are important similarities with Konstan's findings: Both studies identified a steeper rate of decline in lung function up to the age of 25, and both studies identified fewer significant covariates, including sex and *P. aeruginosa* effects, as compared to the comparable analyses in the <18 age group. These findings may be due to lower sample size in the adult population, and survivor effects, or they may be a true reflection of shifting risk factor effects with age. Another possibility is that risk factors that have been present for many years may no longer have a demonstrable effect on lung disease. It is, however, noteworthy that there were still demonstrable sex effects in adulthood in both studies.

Studies like those undertaken by Konstan et al, and study 1 in this thesis, using longitudinal data on large numbers of individuals and mixed-effects model approaches, are ideal to delineate population level risk factors, and are important for a number of reasons. They allow clinicians to better understand, and predict the likely lung function trajectories of populations with similar characteristics, and thus allow identification of populations with multiple risk factors for lung function decline where a higher degree of vigilance might be appropriate. Furthermore there are implications for trial design, and assessment of new therapies, in that these results help triallists to identify populations with higher rates of lung function decline in which therapies may be studied in order to increase the chances of demonstrating significant effects. These studies also highlight the complexity of studying the potentially shifting effect of risk factors for lung function decline in populations.

We can compare and contrast the insights gleaned from the UK analysis (study 1) to those from the Danish analysis of %FEV<sub>1</sub> decline in study 3. The UK analysis, along with the Konstan et al studies discussed above, are adequately powered to assess the effect of covariates on between-individual differences in lung function trajectories, since these datasets contain a large number of individuals, with relatively short

longitudinal traces. By contrast the Danish analysis featured very long, and frequently measured %FEV<sub>1</sub> trajectories, on about 500 individuals, and as described in study 3 this required a novel modelling approach. The Danish dataset thus provided the opportunity to study within-individual changes in %FEV<sub>1</sub> in great detail, whereas it was less well powered to study between individual differences.

Population averaged %FEV<sub>1</sub> decline in the Danish population was significantly influenced by age, *Pseudomonas* colonization status, pancreatic function, birth cohort and CFRD (Taylor-Robinson et al., 2012a). These risk factors have been identified in other studies, including the Konstan studies, and study 1 of the UK population, and the PA effect was similar in Denmark to that found in the ECFS population (Konstan et al., 2007a, Konstan et al., 2012, Taylor-Robinson et al., 2012a, Taylor-Robinson et al., 2013a). However, the Danish analysis in study 3 suggested linear population average covariate effects over time, as opposed to the changing effects demonstrated in the US studies. This is likely to be a reflection of the small cross-sectional sample size in Denmark, as further suggested by the lack of a demonstrable sex or genotype effect.

The key insights from the Danish analysis arise from the description of the variogram for %FEV<sub>1</sub> decline, which has never been done before. This was largely the result of serendipity on my part. The initial rationale for the Danish study was the availability of individual level SES data through the Danish registers, and it was a fortunate surprise to come across such a rich dataset that had never been analysed longitudinally, in combination with the expertise and support of my supervisor Peter Diggle, who has developed many of the methods to analyse data of this nature.

*The key findings from the Danish analysis, with regard to the clinical epidemiology of CF are:*

- *The large error in repeated measurements of %FEV<sub>1</sub> within individuals (average within person SD of 6 percentage points).*
- *The long-term, but exponentially decreasing correlation in %FEV<sub>1</sub> measures over time (up to 15 years), quantified in the variogram.*

These findings have implications for the interpretation of abrupt changes in lung function at the individual level in CF, much of which may be due to recoverable short-term fluctuation, and measurement error. There are also implications for longitudinal study design, and RCTs with %FEV<sub>1</sub> as the primary outcome, where separating a treatment signal from the noise is clearly a challenge, and requires large sample sizes.

The findings of study 3 also raise questions about the plausibility of *individual* level prediction models for %FEV<sub>1</sub> decline, like those suggested by Konstan et al (Konstan et al., 2007a, Vandevanter et al., 2010) which purport to assist prediction at the individual level. For instance, Konstan's paper (Konstan et al., 2007a) suggests that one may be able to use the linear combination of fixed-effect covariates (risk factors for lung function decline) for an individual to predict the likely lung function trajectory over time. This does not, however, take into account the other components of variation – the measurement error, and the within person correlation over time (random effects), which the analyses presented in this thesis have suggested are important, and have an important bearing on the prospects for individual prediction. Furthermore in the Vandevanter analysis that outlines the development of a 'pulmonary outcome prediction score' (POP) system the abstract states (Vandevanter et al., 2010):

*“These simple integer-based POP algorithms employ variables available at clinic visits and can be used to predict the probability of different future pulmonary outcomes for individual patients and patient populations.”*

Further research is required to understand the value of clinical tools such as these at the individual level, and researchers should be careful not to conflate individual and population level prediction.

%FEV<sub>1</sub> remains an influential driver for treatment of pulmonary exacerbations in CF (Rabin et al., 2004), and the key recommended outcome in clinical trials for new therapies (European Medicines Agency, 2009). Clearly it would be ideal if clinicians were able to predict the future for their patients on the basis of data routinely collected in clinic. However, the reality appears more complicated than this, and

some have begun to question the value of %FEV<sub>1</sub> as the key clinical outcome measure of interest in CF. Rosenthal's entertaining piece starts thus (Rosenthal, 2009):

*“Whenever I see a CF doctor, my mind forever superimposes the scene from the film The Producers where Gene Wilder's Leo Bloom clutches his blue blanket but in this case the blanket has FEV<sub>1</sub> written on it – we are obsessed with it.”*

*(see <http://www.youtube.com/watch?v=PTGZKmdfdzI>)*

Rosenthal illustrates a serious point, however, and the Danish analysis presented in this thesis adds to this debate, and suggests some way forward: The modelling framework provides a more realistic estimate of the underlying lung-function trajectory of people with CF, by acknowledging both the imprecision in individual measurements over time and the correlation structure of repeated measurements on the same individual, issues that have all too often been disregarded in the past (Taylor-Robinson et al., 2012a, Taylor-Robinson et al., 2013b).

### **Understanding longitudinal risk of chronic *P.aeruginosa* colonisation in CF**

Study 1 provides, for the first time, a multivariable longitudinal model for the risk of *P. aeruginosa* colonisation over time in the UK CF population. As well as demonstrating an association with social deprivation, this model describes the time-dependent nature of *P. aeruginosa* acquisition, and suggests that female sex, homozygote delta F508 status, CFRD, PI, lower %FEV<sub>1</sub>, and not having been diagnosed through new-born screening are associated with an increased risk of chronic *P. aeruginosa* acquisition. Study 1 appears to be the largest longitudinal study to date of factors influencing *P. aeruginosa* colonization, and the only study at a population level (Kerem et al., 1990b, Kosorok et al., 1998, Maselli et al., 2003) (Rosenfeld et al., 2012a). Schechter et al's cross-sectional study of the effect of SES on *P. aeruginosa* in the US showed Medicaid insured patients were more likely to have *P. aeruginosa* infection than were patients who were not eligible for Medicaid insurance, but when adjusted for %FEV<sub>1</sub> there was no statistically significant difference (Schechter et al., 2001).

*P. aeruginosa* is clearly associated with worse outcomes for people with CF, in terms of survival, lung function, pulmonary exacerbations and nutritional status (Emerson et al., 2002, Konstan et al., 2007a, Rosenfeld et al., 2012a), and the association with worse growth and lung function have been corroborated in study 1 as described above (Taylor-Robinson et al., 2013a). Further research is required to understand the interaction of SES, increased time in hospital, and risk of *P. aeruginosa* acquisition in the UK population, and to identify modifiable risk factors. Rosenfeld et al's recent study in a US cohort (Rosenfeld et al., 2012a) did not identify any modifiable risk factors that were associated with decreased age at *P. aeruginosa* colonization, including exposure to second-hand tobacco smoke, breastfeeding, day-care attendance, hot tubs, or wood burning stoves. They suggest a possible association of better nutritional status with delayed acquisition of *P. aeruginosa*, also demonstrated by Ranganathan et al (Ranganathan et al., 2011). These relationships could fruitfully be explored in further studies of the UK registry.

### **Understanding the sex effect in CF**

The analysis of the UK CF registry presented here has, for the first time in a UK-wide cohort, demonstrated worse outcomes for females from the outset, in terms of lower weight, height and BMI SD scores in the early years; increased rate of decline of %FEV<sub>1</sub> from age five; and increased prevalence of *P. aeruginosa* colonisation in childhood and adulthood. These associations were independent of factors such as SES, and the association between female sex and higher prevalence of *P. aeruginosa* colonisation was not modified by adjustment for %FEV<sub>1</sub>, suggesting that female sex may have an independent effect on risk of *P. aeruginosa* acquisition. Study 1 did demonstrate a period of increased rate of decline of lung function in men in the early adult period, corroborated by Konstan's study in the US (Konstan et al., 2012).

Female sex has now been identified as a negative prognostic factor in CF in several countries, registries, and CF care centres (Buzzetti et al., 2009, Salvatore et al., 2012). The effect of sex on morbidity and mortality in CF, with females having worse outcomes, has been a common finding in large epidemiological studies, first suggested in US centres (Kerem et al., 1992, Kosorok et al., 1996), and then

confirmed in US population level registry studies (Rosenfeld et al., 1997, O'Connor et al., 2002). There have been similar findings in the UK (Dodge et al., 2007), with Barr et al in the UK (Barr HL, 2011) suggesting that despite overall improved survival in the 21st century, females continue to be more likely to die below the median age of death compared to males, a pattern that has persisted since the 1960s. There has been recent debate about the sex gap, suggesting that this may be narrowing over time as a result of improving treatment (Verma et al., 2005, Viviani et al., 2011). In terms of use of health services, a large study in Canada has demonstrated increased risk of hospitalisation in females (Stephenson et al., 2011).

The cause of the sex gap remains unclear, however. Some studies suggest that females may be more likely to become colonized with *P. aeruginosa* at an earlier age (Kerem et al., 1990b, Demko et al., 1995, Maselli et al., 2003, Levy et al., 2008, Rosenfeld et al., 2012a), and this may be related to the effect of oestrogen (Zeitlin, 2008, Chotirmall et al., 2012). Other biological reasons have been suggested, including increased occurrence of CFRD (Barr HL, 2011). Social explanations have also been suggested. These may relate to gender roles, such as a possible propensity to less exercise in childhood in girls, and an increased tolerance of poor nutritional status in adolescent girls with CF, fuelled by the societal pressure to appear thin (Schechter, 2003, Schechter, 2004).

Study 1 shows that the sex effect on outcomes is clearly apparent in the UK CF population. Future studies could usefully explore mediating pathways that may help explain the gap, focussing on the relationship between early nutrition, *P. aeruginosa*, and %FEV1. Furthermore, it may also be possible to test the hypothesis that the sex gap is reducing in successive cohorts. This was not pursued in the UK analysis, where the cohort effect was treated as a nuisance factor in the analysis.

### **Understanding the effect of newborn screening on CF outcomes in the UK**

In the UK population, in the analysis presented in this thesis, screening was associated with improved weight and height for age, reduced risk of *P. aeruginosa* colonization, and a shallower rate of decline in lung function in the <18 analysis (Taylor-Robinson et al., 2013a). The study of the UK registry presented here builds on the papers from Sims et al that were amongst the first publications to use the data

from the UK CF database, which suggested a reduction in morbidity, treatment requirements and healthcare costs in screened patients (Sims et al., 2005a, Sims et al., 2005b, Sims et al., 2007a, Sims et al., 2007b). Furthermore these findings support the growing body of evidence that newborn screening improves outcomes in CF (Salvatore et al., 2010, Greasemann and Ratjen, 2013), and the results specifically corroborate the findings of the systematic review of RCTs of neonatal screening in CF which shows less malnutrition in infants diagnosed by neonatal screening (Southern et al., 2009).

The finding of reduced prevalence of *P. aeruginosa* colonization in the UK screened population is important, since the Wisconsin Neonatal Screening Trial raised concerns that some of the benefits of screening may be offset by an increased risk of earlier *P. aeruginosa* acquisition (Kosorok et al., 1998), although this has not been supported in more recent studies (Wang et al., 2001, Sims et al., 2005b, Baussano et al., 2006, Rosenfeld et al., 2012a).

The UK national CF screening programme offers rich opportunities for further research in combination with the registry. The screened individuals included in the analysis presented here (n = 1309), are mainly those individuals identified by the regional screening programmes for CF that were in place prior to the national newborn screening programme. In England, all babies delivered in the East Anglia, Northampton, Trent and Leeds districts were screened for CF prior to the national roll-out (Pollitt et al., 1997, Southern, 2004, Sims et al., 2005a, Sims et al., 2005b).

From the analysis presented in study 1 we can speculate that new-born screening may be associated with a narrowing of health inequalities in early weight gain, and this will be a testable hypothesis in a few years time once there are data on weight gain in the first three to four years of life in the screened population. Furthermore, studies of the incident screened population in the UK population offer a unique opportunity to study the mechanisms behind the early evolution of differential outcomes in CF, and their consequences over the life-course, in a homogeneous population at the outset. This will lead to a 'cleaner' dataset and population than the one studied in the analysis here, without the complication of survival bias, and any effects associated with missing data prior to diagnosis. The growing body of evidence around the evolution of early lung disease (Greasemann and Ratjen, 2013),

typified by the ARREST CF studies in Australia (Sly et al., 2009, Stick, 2009, Stick et al., 2009, Mott et al., 2012), has been dependent on these features of the CF population identified through new-born screening. I can see great potential for data linkage between the UK registry and smaller clinical studies of this type. Specially collected clinical data could be used to augment the data collected in the registry, and the registry could facilitate the long-term follow up of patients studied in clinical trials of early patho-physiological changes in infants and children with CF. Studies of this nature will be critical to assess the utility of early interventions, implemented before any lung disease is established, that have the potential to deliver long-term benefits for patients with CF (Greasemann and Ratjen, 2013).

## **Contribution to the knowledge base around health inequalities**

Using CF as a case-study to investigate pathways to health inequalities provides a rare opportunity to investigate how the interaction of social, genetic, and healthcare factors leads to differential outcomes, in a disease that does not have a socio-economic bias at the outset. This has led to a number of key findings that contribute to our knowledge base around health inequalities more generally, which are discussed below. The WHO Commission on Social Determinants of Health, chaired by Sir Michael Marmot, emphasised the urgent need for a stronger focus on this type of research. In order to develop effective interventions, the report states that it is essential to measure the problem, expand the knowledge base and build research capacity (CSDH, 2008). The studies in this thesis add to the evidence base, and through undertaking this work I have developed key skills that build research capacity in this vital area.

## **Social inequalities, even for genetic diseases**

The findings presented in this thesis pose a challenge to those who might wish to invoke narrow genetic or biological explanations for inequalities in health. The genetic argument for health inequalities suggests that inequalities stem from the social distribution of genetic material responsible for causing illness, to the disadvantage of those further down the social strata, but there is little empirical evidence to support this (Holtzman, 2002).

CF is the archetypal classically inherited genetic disease that I first learnt about at school, caused by a single gene defect that is both a necessary and sufficient cause of a chronic illness. CF was one of the first diseases for which the precise genetic mechanism was elucidated, with the sequencing of the CFTR gene, and yet knowledge of the genetic mutation does not predict clinical outcome (Kerem et al., 1990a, 1993, Koch et al., 2001, Schechter, 2004). Study 1 demonstrates an even distribution of incident cases in CF. Study 1 also shows a slight difference in the distribution of delta F508 alleles by SES, with a shallow trend towards fewer heterozygote delta F508 carriers and more people with no delta F508 genes with increasing level of deprivation. An explanation for this may be the greater proportion of non-white patients in more deprived groups.

The variability in clinical presentation among patients with identical genotypes is presumably due to a multitude of factors, both genetic and non-genetic, and there is currently great interest in the role of modifier genes in CF. This has led to a number of recent studies into modifier genes that may influence the effect of the CFTR mutation (Wolfenden and Schechter, 2009, Cutting, 2010), most notably the TGF $\beta$ 1 gene where there seems to be consistent evidence to support its association with poorer lung function (Drumm et al., 2005). The interaction between the TGF gene and ETS exposure has been mentioned previously, and is an area for further research (Collaco et al., 2008). Despite the genetic origin of CF, in this thesis I have shown that outcomes in CF are socially patterned, demonstrating that social factors are leading to profound differences in the course of the illness, which cannot be explained by genetic differences.

## **Inequalities from the start**

The finding that inequalities start early in the life course and then track through until later life, even for a genetic disease like CF, supports the growing evidence base around the early origins of health inequalities. The convergence between the inequalities literature, and the current direction of CF research is also striking, whereby both disciplines are suggesting that the early years are critical. In their recent piece on early lung disease, for instance, Grasmann et al state (Grasmann and Ratjen, 2013):

*“The infant and preschool age could represent a unique period of opportunity to postpone or even prevent the onset of cystic fibrosis lung disease.”*

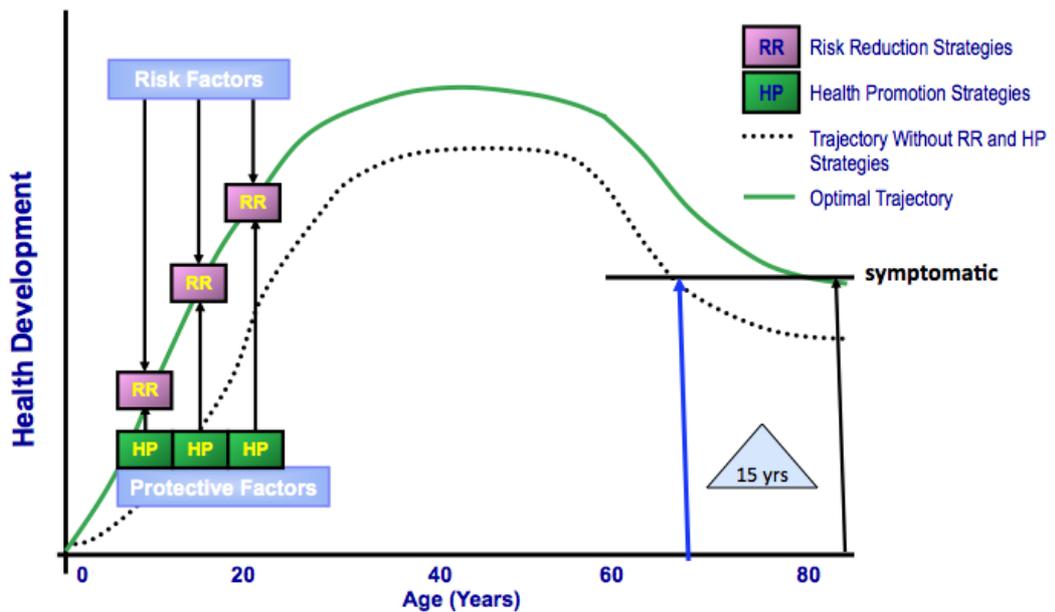
Compare this to possibly the key recommendation of the UK Marmot review (Marmot et al., 2010b):

*“Action to reduce health inequalities must start before birth and be followed through the life of the child. Only then can the close links between early disadvantage and poor outcomes throughout life be broken.”*

The findings of this thesis support the growing consensus that early disadvantage tracks forward, to influence adult health in later life, and that children who start behind tend to stay behind (Kuh et al., 2004, Galobardes et al., 2008, Marmot et al., 2010b). This has important implications for public health policy, and suggests that children may have optimum or sub-optimum trajectories on the basis of early life experiences. As Figure 80 below illustrates, these early life trajectories can determine the point at which an individual becomes symptomatic for a particular adult chronic disease. Thus interventions targeted in the early years may thus delay the onset of limiting illness in later life.

**Figure 80: How risk reduction and health promotion strategies influence health development**

*Adapted from (Halfon et al., 2000)*



The Y-axis in Figure 80 represents some general phenomena relating to health development, but it is striking that this plot is increasingly cropping-up in the life-course literature, with different measures of health on the Y-axis. For instance, this model has been applied to the development of lung health (Kuh et al., 2004), mental health capacity (Kirkwood et al., 2008), and features in a recent Lancet paper in a very similar form with ‘behavioural competence’ on the Y-axis (Walker et al., 2011). I consider the policy implications of this – the need for increased investment in the early years – in more detail later in this chapter.

**Health services and health inequalities**

We have discussed the variations in recorded use of treatments by SES for people with CF in the NHS. What do the studies in this thesis, using CF as a tracer condition, tell us more broadly about the role of health services in the fight against health inequalities?

An important starting point is to understand what equity in healthcare might look like? Restating Margaret Whitehead's definition:

- equal access to available care for equal need;
- equal utilization for equal need;
- equal quality of care for all (Whitehead, 1990).

Need, in the context of the definition above, can be taken to relate not just to disease severity, but also to disadvantage. On this basis one can argue that equity is being achieved for certain aspects of CF care in the NHS, since the data suggests that clinicians are responding to both disease severity, and level of disadvantage when making treatment decisions.

The Commission on Social Determinants of Health highlights that inequitable delivery of health care is an important factor in the generation of health inequalities (CSDH, 2008). Whilst health services are generally not the cause of health inequalities (Marmot et al., 2010b), they can play an important role in the amplification and perpetuation of disadvantage in a number of ways, for example through the so-called 'inverse care law' (Hart, 1971) or the 'medical poverty trap' in health care systems that are not free at the point of access (Whitehead et al., 2001). The famous Inverse Care Law, described by the British GP Tudor Hart (Hart, 1971) states:

*“The availability of good medical care tends to vary inversely with the need for it in the population served. This inverse care law operates more completely where medical care is exposed to market forces, and less so where such exposure is reduced.”*

This describes the situation whereby those who most need high quality health care, end up receiving less and/or lower quality care. For this reason, striving for equity in health service delivery, as defined by equal use for equal need (Whitehead and Dahlgren, 2007), is a central principle for health care systems (Whitehead, 1992), including in the NHS.

In the UK, the Darzi report noted the relationship between low SES and poor health and suggested that the NHS has a pivotal role in providing excellent services as “a matter of fairness” (Darzi, 2008). The controversial new Health and Social Care Act

further enshrines in legislation explicit duties on the Secretary of State, NHS Commissioning Board and clinical commissioning groups (CCGs) to have regard to the need to *reduce inequalities in access to, and to the outcomes of healthcare* (DH, 2012c). Many dissenting voices have suggested the legislation is likely to have significant adverse effects on health inequalities (FPH, 2010, Pollock et al., 2012a, Pollock et al., 2012b). Whilst this may well be the case, the wording of the legislation could serve as a lever to hold those responsible for health services to account in terms of taking action to reduce inequalities in outcomes.

In these troubled times for our NHS, Study 1 provides some evidence that the system can take into account the extra needs of disadvantaged children and their families in the UK to deliver care that may be effective in reducing, or at least limiting any increase in health inequalities. Identifying levelling-up interventions such as this is the ‘holy-grail’ for researchers and policy makers interested in reducing health inequalities (Klasen, 2004), as Margaret Whitehead asserts in a paper for WHO (Whitehead and Dahlgren, 2007):

*“...the only way to narrow the health gap in an equitable way is to bring up the level of health of the groups of people who are worse off to that of the groups who are better off.”*

Study 1 presented here contributes to this evidence base, and provides encouragement that health services can make a difference in the early years. Further research in this area is required to address the so-called ‘inverse evidence law’ around interventions to reduce health inequalities (Milton B, 2011).

Study 1 also sheds some light on potential differences between the delivery of chronic care for children versus that delivered to adults, by demonstrating an increasing propensity to inequality in delivery of both inhaled CF treatments in the adult population, as compared to the paediatric age group. This corroborates the broader literature, in terms of access to and use of care in chronic diseases, which suggests a different picture in adults and children: In adults, a systematic review of universal health care systems by Hanratty et al suggested a decline in need adjusted use with increasing deprivation, especially in use of specialist hospital services, but reasonably equitable access to primary health care (Hanratty et al., 2007b). A more recent European study found large inequalities in the utilisation of specialist care for

chronic diseases, which were not compensated by utilisation of GP services (Stirbu et al., 2011). In study 1 here I demonstrate a mixed picture for adults, with marked pro-poor bias for IV and nutritional therapies, but inequality in the use of inhaled therapies.

This contrasts with analyses of children's use of health services in the UK, where the studies have tended to show increased use with lower SES, and arguably equitable delivery of care in the studies that have adjusted for need, notwithstanding the difficulties of undertaking precise adjustment. Cooper et al found no evidence that children and young people's use of health services varied according to their SES, in an analysis that adjusted for need on the basis of self-reported health status, a rather crude measure, in an analysis of the British General household survey (Cooper et al., 1998). Saxena et al demonstrated increased use of primary care for disorders such as infections, asthma, and injuries and poisonings in children with lower SES (Saxena et al., 1999). In a further study by the same author children's use of primary and secondary health services reflected health status rather than SES, and on this basis the authors suggest that equity of access has been partly achieved (Saxena et al., 2002). A more recent study in the UK demonstrated a clear association between adverse SEP at the time of birth and increased hospital inpatient admissions, days, and costs during the first 10 years of life (Petrou and Kupek, 2005). A study of children in Nordic countries demonstrated equitable access to primary care, but inequality in use of specialist hospital services after careful adjustment for health status on the basis of questionnaire data (Halldorsson et al., 2002).

It remains to be seen whether the increasing inequality observed in use of inhaled treatments in CF in adults reflects general differences between paediatric and adult care for chronic disease, or is specific to differences in the CF model of care, perhaps explained by the potential for paediatricians to act in a more paternalistic manner towards the children in their care, in contrast to the relationship between an adult patient and the CF care team, where adult patients can 'vote with their feet'. Further studies should focus on identifying aspects of CF care in children, and adulthood that may be important in reducing health inequalities.

## **Methodological advances**

### **Using CF as a tracer condition**

The studies in this thesis are the first explicitly to recognise the utility of using an autosomal recessive condition as a tracer condition to study pathways to health inequalities, leading to the general insights described above. Sickle-cell disease, also autosomal recessive in inheritance, would provide another useful case-study, albeit possibly more complicated by the interaction with ethnicity. There have been a number of small studies suggesting social differences in outcomes and health care use in this disease (Okany and Akinyanju, 1993, Ellison and Bauchner, 2007, Hijmans et al., 2010, Animasahun et al., 2011), and there is the potential to develop this line of research further.

### **Novel modelling approach**

Study 3 describes a novel modelling approach for analysing changes in %FEV<sub>1</sub> over time, which can be applied at the individual level to interpret the clinical significance of sudden changes in %FEV<sub>1</sub>, and at the population level to quantify the effect of factors such as *P. aeruginosa* acquisition. This approach extends the commonly used random intercept and slope model. Whilst the statistical framework for this analysis is not new (Diggle et al., 2002), this approach has not been used to analyse lung function data, and is not in widespread use in epidemiology. In this respect study 3 represents an important piece of knowledge transfer, whereby a methodology is introduced to a new disciplinary area.

We can apply our approach to modelling changes in %FEV<sub>1</sub> over long follow-up periods. This is in contrast to the widely used random intercept and slope approach that has been applied in studies of CF and COPD, over short-term (Vestbo et al., 1999, Mastella et al., 2000, Konstan et al., 2007a, MacLean et al., 2008) and longer-term follow up periods (Corey et al., 1997, Hnizdo et al., 2005, Stern et al., 2007, Kohansal et al., 2009). The development and testing of the new approach was facilitated by the nature of the Danish CF register – to my knowledge there are no other datasets that contain such frequent (monthly) measures of lung function on individuals measured over very long periods (up to 31.5 years). However, the fact that the data are from Denmark does not influence the validity of the methods, since

these are essentially context-free. Furthermore, this method does not exploit any features of the data that are unique to CF, and would be equally applicable to other clinical areas that generate long sequences of repeated measurements.

## **Critique of overall study design**

The strengths and limitations of each study were discussed briefly in Chapters 4-6. Below, I explore the strengths and limitations of the overall approach and study design, focussing on the data sources and the methods used to analyse the data.

### **Key strengths of the datasets**

A key strength of the two longitudinal datasets analysed in this thesis is that they are population level datasets, capturing the vast majority of the prevalent population of people with CF in the UK and Danish population. For instance, the UK dataset is estimated to capture over 99% of the UK CF population (Taylor-Robinson et al., 2013a). Furthermore, both datasets are of high quality, with systems for data cleaning and checking. The datasets are both large, in complementary ways. The UK dataset represents one of the largest national CF datasets outside of the US (Buzzetti et al., 2009), and this has provided the power to precisely estimate parameters in the UK analysis, especially with regard to between-individual differences. The Danish dataset, on the other hand, is characterised by extraordinarily frequent follow-up, over long periods of time, and this has allowed precise estimation of within-individual changes, through estimation of the variogram function (Taylor-Robinson et al., 2012a).

The UK Registry contains a wide range of clinical, health care, and social information, allowing for robust adjustment for appropriate covariates in the analyses in study 1. Although the Danish registry contains mainly clinical outcome data, linkage to national datasets in Denmark through the unique Danish ID number allows the data collected within the registry to be supplemented extensively. Patient registries such as these are powerful tools designed specifically to improve our knowledge of disease progression and management. They provide systematically collected data that reveal patterns in disease diagnosis, treatment and outcomes over time (Salvatore et al., 2012), and have provided some of the crucial insights into the epidemiology of CF, as outlined in Chapter 2 (Buzzetti et al., 2009, Salvatore et al., 2010, Salvatore et al., 2011, Salvatore et al., 2012).

The findings from the studies contained within this thesis thus provide information pertinent to the whole population of either the UK or Denmark, and this is one of the

great strengths of registry analyses. This is in contrast to studies that are conducted at a care-centre level, the findings of which can only be cautiously generalized (Stanton et al., 2011, Taylor-Robinson et al., 2011). For example, with regard to CF, study 1 is the first to demonstrate a clear effect of SES on *P. aeruginosa* acquisition at a population level, with all of the previous studies having been undertaken in centre-based or sub-national populations (Kerem et al., 1990b, Kosorok et al., 1998, Maselli et al., 2003, Rosenfeld et al., 2012a). The findings from the studies in this thesis cannot therefore be easily dismissed as artefacts, due to quirks of data collection or practice particular to a specific centre.

A further consequence of the high level of population coverage in the UK, coupled with a universal health care system, is that the analyses in study 1 cover individuals across the full range of the socio-economic spectrum in the UK. Concerns about generalizability of findings from the US have been discussed above, but a further issue regarding data on SES outcomes in the US is the extent to which studies capture the most deprived people in the population, and the extent to which there is any selection bias into the US registry (Gliklich and Dreyer, 2010). The studies on SES and outcomes in CF from the US pre-date President Obama's historic affordable care act (Jaffe, 2012), and were undertaken during a period when the number of uninsured people in the US was up to 50 million people. The profile of uninsured people in the US has changed little over time, being largely the poor, ethnic minorities, new immigrants, the poorly educated, and those in poor health (Hoffman and Paradise, 2008). Furthermore, concerns were raised at the North American CF conference 2012 regarding access to care for children of 'unregistered' parents (illegal immigrants), whose parents could be subject to deportation if they follow the social-security proceedings necessary to secure access to Medicaid treatment for their children (Sufian, 2012). In such a context one can speculate that CF registry analyses in the US are more likely to miss the most disadvantaged individuals, and thus suffer more from selection bias, as compared to the UK.

### **Key limitations of the datasets**

Registries can provide valuable insights into variations in clinical outcomes, quality of care, and the safety and/or effectiveness of treatments. In this respect they are powerful tools for research and quality improvement in the context of complex and

resource-intensive chronic diseases, such as CF. However, the usefulness and applicability of registry data relies on the quality of the data analysis, and appropriate interpretation, bearing in mind the shortcomings of the retrospective, routinely collected data (Gliklich and Dreyer, 2010, Salvatore et al., 2012).

Both the UK and the Danish datasets are subject to left truncation, whereby the datasets capture the prevalent population at the inception of data-collection, and incident cases subsequent to this. This leads to potential survivor bias, whereby the prevalent population at the outset of data collection represent healthier individuals from their respective birth cohorts who have survived to the point of being included in the dataset. This is a common issue in registry analyses, and is a recognised feature of previous analyses in the US (O'Connor et al., 2002) and Denmark (Frederiksen et al., 1996). This has been particularly highlighted as source of bias in many of the studies that have estimated survival in CF, especially those using age-specific survival rates (Corey, 1996, Frederiksen et al., 1996, Lewis, 1998). This was one reason for focussing the UK analysis in Study 1 on longitudinal clinical outcomes, rather than survival. There is the recognition that cohort effects in the dataset are likely to represent a mixture of survivorship effects, and the 'true' cohort effects representing improving treatment over time, as demonstrated in other studies (Buzzetti et al., 2009). Thus cohort effects were adjusted for carefully on the basis of year of birth in the main analyses, but these were treated as nuisance variables, which need to be adjusted for in order to identify independent SES effects.

Furthermore, as a result of left-truncation we can speculate that the true SES effect on outcomes may potentially be underestimated in studies 1 and 2, due to failure of the registry to capture potentially sicker patients who died prematurely compared to their contemporaries. Whilst the longitudinal models used adjust for dropout due to death in individuals captured in the dataset, we could do nothing about any bias due to deaths prior to inception of the registry, other than acknowledge the issue. With longer follow-up as the UK registry matures, separating age and cohort effects will become possible, and eventually it will be possible to analyse incident individuals alone, ensuring that the longitudinal experience of all individuals from a particular birth cohort will be captured (Gliklich and Dreyer, 2010).

Although the UK registry contains a large amount of clinical information, important putative mediators of SES effects are not collected in the analysis, as previously discussed with reference to Figure 78. Data on exposure to smoking is the most notable. Ideally, data should be collected on exposure to smoke exposure *in utero*, coupled with data on birth weight, and consequent capture of exposure to parental smoking in the home environment. Notwithstanding the potential difficulty of asking questions about parental smoking sensitively, and the possibility of incorrect responses, it is a serious omission that these data are not collected in the UK Registry, or in the US and Denmark, in the core dataset for a chronic respiratory illness. Collecting such data would be a first step to quantifying the problem, and for organising appropriate referral for parents who would like help to quit smoking.

Approximately 90% of individuals had a valid postcode. Some postcodes were invalid, predominantly due to incomplete or incorrect entry of postcode information, which meant that these could not be linked to area deprivation scores. This is unlikely to be a systematic process causing bias. However, since many postcodes were captured during the transition to the electronic data system, people who died prior to this are under-represented in the dataset, with improved postcode coverage evident in more recent cohorts. This is a potential source of bias, and we can speculate that this may lead to underestimation of any social gradient in outcomes, if more disadvantaged individuals are more likely to die and thus be missing. The even distribution of the overall population by SES, however, is reassuring, however, and means that I was able to estimate SES effects on the repeated longitudinal outcomes across the whole socio-economic spectrum. Furthermore, restricting the analysis to the period after 2005 did not materially alter the SES or other covariate effects in sensitivity analysis. However, the disproportionately missing postcode data on people that died poses significant challenges for conventional survival analysis. This represents an unfortunate omission in the overall picture presented in this thesis, meaning that I was unable to corroborate the findings of Britton (Britton, 1989) and Barr (Barr HL, 2011), who showed social gradients in CF mortality, using death certification data.

Missing data on outcomes of interest is a perennial problem with longitudinal studies, but in study 1, after application of the eligibility criteria, there were very little missing data in the final analysis, since there is good coverage of the basic

demographic variables in the analyses, and clinical outcomes such as weight or %FEV<sub>1</sub> are key variables for collection at annual review. Furthermore, as discussed previously, the modelling approach I used handles missing data under the assumption of MAR, given the covariates in the regression. Fitting the model by maximum likelihood thus automatically corrects for selection bias that depends on a patient's observed measurements prior to death, although not for any additional dependence on unmeasured features (Diggle et al., 2002, Fitzmaurice, 2004).

Lack of data on other measures of individual level SES in the registry is an important limitation. The UK registry, since the transfer to Port CF, does collect data on both maternal and paternal educational level, but the coverage of this question was very poor (for example, approximately 5% of individuals <18 with data on maternal education). Corroborating the SES effects demonstrated at area-level with individual level measures would have improved the analysis of the UK data. Speaking to clinicians about the missing data in this field, some suggested that this question was difficult to ask, and risked irritating patients and parents, since it was not obviously related to clinical care. Collection of individual level socio-economic data, for the explicit purpose of measuring SES, has to be undertaken sensitively, since as Bourdieu (Bourdieu, 1984) and Wilkinson (Wilkinson and Pickett, 2010) note, we are all acutely sensitive to being 'sized up' by other people. Effective ways emphasising the importance of collection of data such as this should be explored further.

### **Key strengths of the analysis**

The analyses in this thesis have used the modern longitudinal data analysis techniques described earlier in the methods section. These approaches have been strongly recommended to study changes in lung function in patients with CF (Edwards, 2000), and allow one to distinguish cross-sectional effects between individuals and longitudinal age related effects within individuals. Furthermore, they permit us to model either binary or continuous outcomes, and account for the correlation between repeated measurements from the same individuals. In this thesis, these approaches have been applied for the first time to the large national datasets described above, leading to the novel findings described at the outset of this chapter. In addition, the data analysis was informed by Professor Diderichsen's (a collaborator) conceptual framework of the pathways from social context to health outcomes, which was employed by the WHO Commission for Social Determinants of Health. This is in contrast to previous analyses of SES and outcomes in CF that have not been explicitly informed by a theoretical perspective.

### **Key limitations of the analysis**

All of the findings in this thesis are based on observational data, and there is thus potential for confounding due to omission of unmeasured variables, and residual confounding due to imprecise measurement of covariates included in the models. With observational studies of this nature one can always argue for the addition of a particular covariate in a model after the event. To guard against this the analyses in this thesis were based on a clear modelling strategy, which generally involved fitting a baseline model adjusting for factors unlikely to be in the causal chain between SES and the outcomes of interest (e.g. age, sex, birth cohort, and genotype). Then deprivation measures were added to the model in question, and finally subsequent models were fitted to explore the influence of other covariates of potential interest.

The studies in the UK used an area-based measure as a proxy for individual SES, and the main analysis and conclusions were based on it. Ideally, the analysis would have employed both individual and area level measures, but no other measures of SES were useable in the UK Registry. The so-called 'ecological fallacy' – ascribing area level associations to individuals – is an important concern when an area level measure is used as the proxy measure for the individual level, and appropriate

caution should be exercised when interpreting and generalising the findings (Schwartz, 1994). However, this is partly addressed by being specific about the level at which the claim of association is being made. I have demonstrated associations at the small-area level, and cautiously suggest that these associations can be taken as proxies for individual level effects.

Area based deprivation measures are commonly used as proxies for individual SES, especially in the UK (Galobardes et al., 2007) and the US, where their use in public health research is widespread (Krieger et al., 2002). It is accepted that this can potentially lead to biased estimates if individual level SES effects are the targets of inference. For instance, the area-level estimate of the association between SES and the outcomes studies could be an underestimate of the individual-level effect, because of the misclassification that arises through giving everyone in a small-area the same score. Furthermore, this problem increases with ecological areas of greater size (Smith et al., 1998). However, it is recognized that associations can be biased in both directions, and if there are large area level effects (e.g. pollution) that are independent of any individual level effects, then the area effect will be an overestimate (Geronimus, 2006, Subramanian et al., 2006, Galobardes et al., 2007, Galobardes, 2012).

Whilst accepting the potential for ecological fallacy, this is minimised in the UK analysis, in comparison to the studies undertaken in the US, because the IMD methodology allows much finer small area resolution, down to the level of census lower layer super output areas (LLSOAs). The IMD score is a widely used measure of deprivation of area of residence for epidemiological studies in the UK (Semple et al., 2011, Taylor-Robinson et al., 2011, Bergen et al., 2012), and is used in national, local government, and NHS reports on tracking inequalities in health and in access to health services (DH, 2012b). Each small area contains about 1500 people, and in this respect the IMD allows much finer resolution than the US analyses (O'Connor et al., 2003, Schechter et al., 2009, Schechter et al., 2011) that have used ZIP code linked income data, since each ZIP code contains about 30,000 people (Krieger et al., 2002). Furthermore, it is reassuring that similar associations have been found in the US studies that use both area and individual measures of SES.

The lack of multiple measures of SES, whether at area level, or individual level, is also a clear limitation. The most commonly used measures of SES relate to occupation, education and income (Mackenbach et al., 1997). This is because each of these measures may potentially capture a different dimension of the complex construct that is SES. Furthermore, in Schechter's more recent studies of health service use in CF in the US he was able to use maternal education, zip code linked area-based income, and Medicaid status (Schechter et al., 2009, Schechter et al., 2011). The general principle that emerges is that multiple measures of SES enrich the analysis, and allow further insight into potential mechanisms. For instance, parental education may be a particularly important measure, reflecting parental disease management skills, independent of income.

A further potential limitation regarding the measurement of SES is the approach taken to combining the deprivation scores from the four UK countries. There are a number of differences between the IMD that make combining them in a robust manner or making meaningful comparisons across the UK difficult (Payne and Abel, 2012). In order to combine the IMD measures from England, Scotland, Wales, and Northern Ireland, I explored standardising these within each country, but this made little difference in the analysis, so the unadjusted raw scores were used. This may have led to some misclassification. Reassuringly, restricting the analysis to English data only made little difference to the effect estimates. Recent guidance has been developed for researchers wishing to combine deprivation scores across the four countries of the UK, and this approach could be applied in subsequent studies (Payne and Abel, 2012).

## **What are the implications for policy and clinicians?**

Michael Schechter, the author of the main studies in the US on inequalities in CF comments thus (Schechter, 2013):

*“The effect of socioeconomic status (SES) on health is well established in the general public, and has been described in both the US and UK cystic fibrosis populations. In The Lancet Respiratory Medicine, David Taylor-Robinson and colleagues present a sophisticated longitudinal analysis of data from the UK Cystic Fibrosis Registry, assessing and confirming this association using small-area deprivation scores assigned by postcode.... In view of the fact that there is now ample evidence of an important SES-related health gradient in cystic fibrosis populations, at least in the USA and the UK, the logical next step is to develop approaches to alleviate it.”*

I wholeheartedly agree that although there is much scope for further research in this area, now is the time for action. One of the strengths of the Diderichsen model, employed in this thesis, is that it provides a framework for considering appropriate policy responses (Diderichsen et al., 2001), and one can think of these at the level of the individual clinician treating a patient with CF, or at the broader public health and policy level.

### Individual clinically focussed actions

For individual clinicians, low SES needs to be considered as a major risk factor for poor outcomes in CF, and appropriate responses to remediate against the adverse effects of low SES should be developed. An important first step is to acknowledge and measure the problem. Without on-going data collection to monitor health inequalities, there can be no policy response. The pioneers of health inequalities research, the likes of Dr Duncan, or Margaret Whitehead, began by demonstrating the extent of the problem. In terms of measuring the problem, the work undertaken in this thesis has been recognised by the UK CF Trust, who are now looking into how registry data could be routinely presented in a way to better understand and monitor the effects of SES. In their latest Registry report they state (CF Trust, 2013a):

*“It is clear to us that outcomes in terms of lung function can be affected by many factors including those related to care (such as use of CF therapies) and those related to the person with CF (such as genotype, age and socioeconomic status). We would like to be able to provide better comparisons of centres so that we can learn from each other and ensure optimum care for the community... We will expect to see changes in the annual report over the next three years as a result...”*

Furthermore, the work presented in this thesis has been highlighted as part of the new UK CF Trust research strategy, in order to emphasise the utility of registry studies to inform best practice (CF Trust, 2013b). Thus the results of the studies in this thesis should go some way to raising awareness, and initiating debates about appropriate responses.

There are a number of steps that could be taken to influence differential exposures. One obvious target for action is to protect newly diagnosed children from ETS, since this may be the single most important explanatory factor for SES-related inequalities in this disease (Taylor-Robinson and Schechter, 2011). Early identification of family members who smoke, collection of that data in the registry, and appropriate counselling and referral to smoking cessation services would be an effective intervention for all patients regardless of social position. Study 1 also demonstrated differential exposure to *P. aeruginosa* acquisition, and this requires further investigation. One concern is that this may relate to hospital-acquired acquisition of *P. aeruginosa*, from other patients with CF, as a result of more time spent in hospital. An appropriate policy response might be to ensure that mechanisms are put in place to facilitate equitable delivery of treatments at home, wherever possible. Home care has been shown to be as effective as hospital treatment for some patients (van Aalderen et al., 1995), but not others (Bosworth and Nielson, 1997, Thornton et al., 2005). It is also less expensive and may be associated with improvement in quality of life (van Aalderen et al., 1995). Access to appropriate home care may also mean that children are less likely to miss school, which is important in terms of reducing the differential consequences of ill health.

There are some further steps that individual clinicians and teams could take to reduce children's exposure to poverty, and its consequences. Whilst beyond the scope of full discussion here, an increased focus on a whole family approach to the care of the child with CF, with appropriate involvement of the full range of social services support available to children and families living in disadvantaged circumstances may help to mitigate some of the effects of low SES. This would include supporting parents to access all the benefits and services that they are entitled to, and working to reduce any perceived stigma associated with using these services. Support with the additional costs of childcare, travel to clinic appointments, and any additional medical expenditure would also help reduce the financial burden on the most disadvantaged families. This should be coupled with support to develop patient and family disease self-management skills (Goldman and Smith, 2002, Smith and Goldman, 2010).

Social workers, and other members of the multidisciplinary CF team have an important role in helping people with CF navigate the welfare system, so that they are clear about the benefits that they are entitled to, and the support available to help people with chronic illness in the workplace. This is particularly important at the moment, in the context of the widespread changes to the welfare and benefits system in the UK, with the drive to reduce the number of people claiming Disability Living Allowance. A recent study by Nash et al demonstrated that the majority of adult patients with CF claim some benefits, and the majority of these were concerned about the planned reforms (Nash et al., 2011). Furthermore, the concerns of the CF community regarding the changes to the welfare system were outlined in a submission by the CF Trust to the Department of Work and Pensions (Department of Work and Pensions, 2011). Further efforts are required to identify effective workplace, rehabilitation, and other interventions to reduce the employment disadvantage experienced by people with CF, particularly for those from more disadvantaged areas.

A key question for practicing physicians addressed in this thesis is what role health care delivery plays in mitigating or potentiating health inequalities in CF. In the UK, I have demonstrated a mixed picture, and action is required to understand the inequalities in access to inhaled therapies uncovered, and to promote the pro-poor delivery of care demonstrated for other treatments, that may be effective in limiting

increases in outcome inequality over time. The potential utility of systematically targeting more intensive therapy at children living in disadvantaged circumstances should be investigated further (Gupta et al., 2009). The adoption of system-based methods to optimize consistency in the use of best care practices might help further minimize variations in prescribed care (Schechter et al., 2009). Furthermore, the early appearance of inequalities, and the potential for decreasing inequality in weight in the first few years of life, focusses policies on the early years, and provides support for new-born screening programmes in CF.

### Broader interventions

Ultimately, however, while individually focussed interventions may be of some limited success, the long term solution to health inequalities in people with CF and in the general population is likely to be one that takes broader action to address the “social determinants of health”(CSDH, 2008). These are the “conditions in which we are born, grow up, work and live”, and include income and income distribution, education, employment and working conditions, housing, food insecurity, race/ethnicity, and gender and gender roles. These factors provide a particularly important context for a family dealing with the stresses of caring for a child with a complex chronic illness like CF over a lifetime. The evidence is clear, unfortunately, that we have made little progress over the last few decades in reducing health inequalities (Marmot et al., 2010b). However, the findings in this thesis present further evidence that the early years represent the key period for targeting interventions to reduce inequalities. An important place to start would be to renew efforts in the UK to reduce children’s exposure to poverty, thus reducing social stratification, by for instance (Marmot et al., 2010b):

- maximising household incomes, by helping parents into employment;
- providing affordable housing;
- providing affordable, high quality child care;
- providing affordable public transport;
- helping families manage debt;
- providing better social security support for families caring for children with chronic illness.

With reference to Figure 78 and Figure 80 above, policy makers should also act to reduce any inequitable distribution of health damaging and health promoting exposures over the course of children's lives. Even achieving an equal reduction in a particular risk factor across the population by SES has the potential to reduce inequalities in outcomes, through the differential vulnerability pathway (Diderichsen et al., 2001). Reducing the consequences of poverty by focussing on child development in the early years is a good place to start. Actions could involve (Field, 2010, Marmot et al., 2010b):

- protecting investment in the early years in the face of budget cuts in the UK;
- shifting expenditure towards the early years wherever possible;
- providing high quality and consistent support and services for parents during pregnancy;
- provision of high quality universal services in childhood;
- supporting families to achieve progressive improvements in early child development, by providing good quality early years education and childcare;
- providing support so that all children can access a healthy diet in the early years;
- providing high quality home visiting services;
- focussing on narrowing the educational attainment gap at all stages.

As health deteriorates, the ability of people to remain in education and employment declines. Being out of work increases the risk of poverty and social exclusion, and is likely to further damage the health of the most disadvantaged. Actions to address differential social consequences could include (Holland et al., 2011a):

- supporting people with chronic illness to find appropriate employment, with a focus on active labour market policies;
- provision of better in-work social security support for people with chronic illness.

## **Conclusions**

CF is the commonest serious inherited disease among Caucasian populations. Intensive support from family and health care services is needed from the time of diagnosis onwards, and most patients die prematurely from respiratory failure. There have been astounding improvements in survival over successive birth cohorts in CF, such that it is estimated that British children born in the 21st century will have a median survival of over 50 years of age (Dodge et al., 2007). However, there remains a great deal of variation in disease progression and survival in CF, much of which is related to social and environmental rather than genetic determinants (Schechter, 2004). Most pointedly, it has been known for over 20 years (Britton, 1989) that people with CF from socio-economically disadvantaged backgrounds die younger than those in more advantaged positions.

CF offers a valuable case for understanding how health inequalities develop. It is an autosomal recessive disease with an asymptomatic (and, until recently, undetectable) carrier state, so unlike many other diseases, SES does not influence who gets CF. SES-related outcome inequalities develop due to the different patterns of exposure to harmful and protective or therapeutic influences over the course of people's lives. Studies from the US and UK show that significant inequalities in key intermediate CF outcomes such as growth and lung function begin early in childhood (Schechter et al., 2001, O'Connor et al., 2003, Taylor-Robinson et al., 2013a) and then persist over time. The early appearance and persistence of inequalities supports the need for interventions that are targeted at the early (and perhaps prenatal) years.

## **Further research**

The studies in this thesis have highlighted a number of areas requiring further research. There are some important issues relating to our understanding of the clinical epidemiology of CF, which are outlined in the sections above. There is much that could be learned through study of the longitudinal clinical outcomes of the increasing number of children in the registry captured through new-born screening, with a focus on how early nutritional development influences lung function and ultimately survival. Any such analyses should consider the effects of SES.

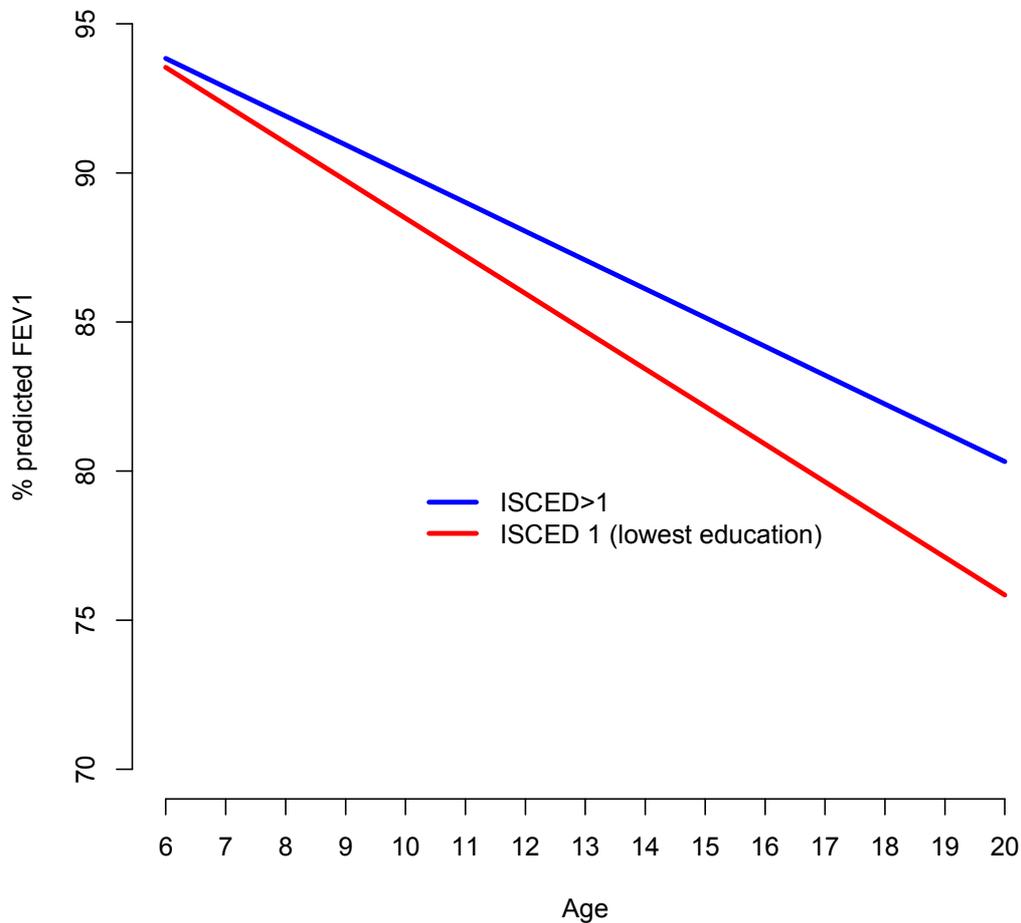
The methods for modelling lung function decline described in study 3 could be usefully applied to longitudinal data collected in other CF registries, to clarify how robust this approach is in terms of predicting changes in %FEV<sub>1</sub> over time, and to better understand how this might inform clinical decision making. Future research could also explore the utility of our proposed model in other diseases such as COPD.

### **Research currently in progress from this thesis**

Analysis of the Danish dataset is on-going. Study 3 in this thesis has outlined the methods for analysing this unique dataset, and the next step is to use this approach in conjunction with data linked through the Danish social registers. Importantly, individual level data linkage in Denmark will allow access to richer individual level socio-economic data than is available in the UK, on patients and their parents. The linked data extends retrospectively to 1981 and contains complete information on an individual's employment, income, healthcare expenditure, hospitalisations, and receipt of welfare benefit. We can use this to explore the effects of a range of individual level parental and patient socio-economic exposures on CF outcomes, and also study changes in socio-economic conditions over time, before and after specific phases of disease development.

At the time of writing I have been able to undertake a preliminary analysis of the effect of parental education level on lung function decline in the Danish population, and this is shown in Figure 81. The figure visualises the same model as that was published in the Thorax paper (Taylor-Robinson et al., 2012a), but with parental education level included, as measured by the International Standard Classification of Education (ISCED) classification. For the purposes of the analysis, this has been dichotomized to compare individuals with the lowest SES, to all the rest, due to the small numbers involved in the Danish analysis. This analysis shows an increased rate of decline of lung function in people with the lowest SES (-0.31, 95% CI 0.06 to 0.56). Further confirmatory analysis using alternative measures of SES, and other CF outcomes (growth, *P. aeruginosa* status, and survival) in Denmark is ongoing.

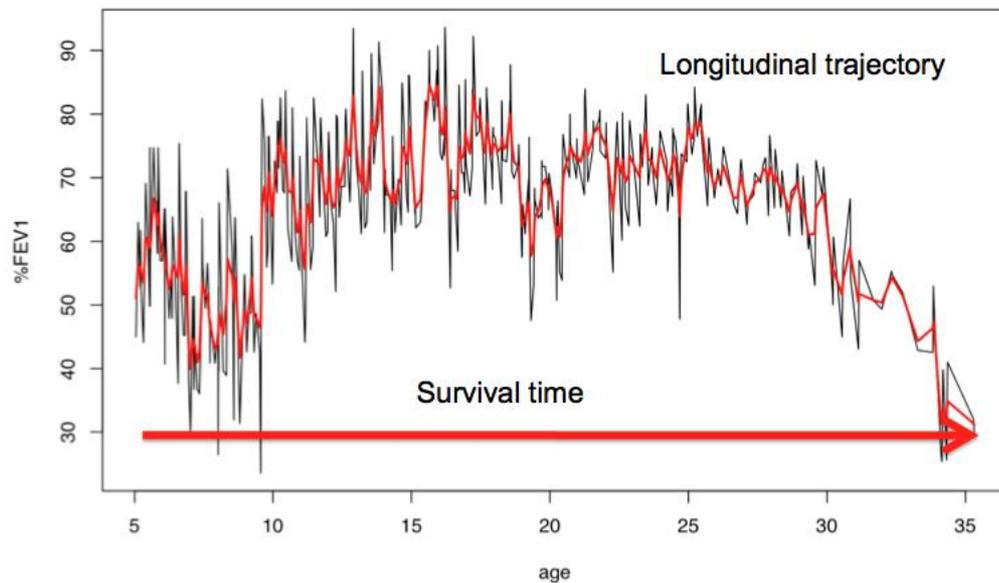
**Figure 81: Effect of parental education level on lung function decline in Denmark**



A further line of on-going enquiry in the UK and Denmark is an exploration of how aspects of an individual CF patient's longitudinal profile of %FEV<sub>1</sub> are related to their survival prognosis. This is intended as an update on the seminal Kerem analysis that first demonstrated the relationship between low lung function and increased mortality (Kerem et al., 1992). The headline figure from the Kerem paper was a 50% risk of death within the next two years, for patients with a %FEV<sub>1</sub> lower than 30%, in the US centre-based sample (Kerem et al., 1992). One can see in Figure 82 below, from the Danish registry, that %FEV<sub>1</sub> drops below 30% on a number of occasions in the first 10 years of life for this individual in Denmark, and we have argued that the error free lung function measurement estimated by the model in Study 3 provides a

better estimate of an individual's true underlying lung function. The intention is now to develop a joint model, which relates aspects of this underlying trajectory to survival chances.

**Figure 82: Joint modelling of lung function and survival outcomes**



In collaboration with biostatisticians, I am using CF data with recently developed methodology for the joint analysis of repeated measurements and time-to-event outcomes. These methods allow us to examine the association between longitudinal changes in %FEV<sub>1</sub> and survival chances, whilst adjusting for, allowing for correlation within patients, trends over time and potentially informative missing values. Key methodological challenges relate to accommodating cohort effects, and biased entry to registry cohorts, as described in the limitations section (Taylor-Robinson et al., 2012b, Barrett et al., 2013).

Finally, building on the work undertaken during the course of this PhD, as part of an MRC Population Health Scientist Fellowship, I have successfully obtained further funding to continue, and extend some of the lines of enquiry developed in this thesis. I intend to explore the effect of SES on the longitudinal risk of asthma, the most common chronic disease of childhood, in a contemporary, representative UK birth cohort called the Millennium Cohort Study. This will involve developing and testing a logic model to explain how inequalities in asthma develop, by quantifying differential exposures to mediating risk factors (e.g. home environment measures,

breast-feeding, smoke exposure), differential outcomes (asthma risk), and potentially differential social consequences (e.g. school performance), by SES, over time.

In this way, I will further develop the longitudinal analysis techniques outlined in the studies in this thesis, in continuing collaboration with experts in this field (Professor P Diggle, University of Liverpool; Dr J Barrett, MRC Biostatistics Unit), to explore the causal relationship between measures of early childhood disadvantage (e.g. child poverty, maternal education), and the development of adverse health outcomes in children. Comparing and contrasting the cases of CF and asthma will further improve our understanding of the social patterning of illness in childhood.

## References

- ADLER, A. I., SHINE, B. S., CHAMNAN, P., HAWORTH, C. S. & BILTON, D. 2008. Genetic determinants and epidemiology of cystic fibrosis-related diabetes: results from a British cohort of children and adults. *Diabetes Care*, 31, 1789-94.
- AKED, J., STEUER, N., LAWLOR, E. & SPRATT, S. 2009. Backing the future; why investing in children is good for us all. *New Economics Foundation, London*.
- ALIGNE, C. A., AUINGER, P., BYRD, R. S. & WEITZMAN, M. 2000. Risk factors for pediatric asthma. Contributions of poverty, race, and urban residence. *Am J Respir Crit Care Med*, 162, 873-7.
- ALLEN, G. 2011. Early intervention: the next steps. *London: HM Government*.
- ALMQVIST, C., PERSHAGEN, G. & WICKMAN, M. 2005. Low socioeconomic status as a risk factor for asthma, rhinitis and sensitization at 4 years in a birth cohort. *Clin Exp Allergy*, 35, 612-8.
- ALTMAN, D. G. & BLAND, J. M. 2007. Missing data. *BMJ*, 334, 424.
- ANIMASAHUN, B. A., TEMIYE, E. O., OGUNKUNLE, O. O., IZUORA, A. N. & NJOKANMA, O. F. 2011. The influence of socioeconomic status on the hemoglobin level and anthropometry of sickle cell anemia patients in steady state at the Lagos University Teaching Hospital. *Niger J Clin Pract*, 14, 422-7.
- ARIAS LLORENTE, R. P., BOUSONO GARCIA, C. & DIAZ MARTIN, J. J. 2008. Treatment compliance in children and adults with cystic fibrosis. *J Cyst Fibros*, 7, 359-67.
- ARMITAGE, P., BERRY, G. & MATTHEWS, J. N. S. 2008. *Statistical methods in medical research*, Wiley-Blackwell.
- ASHTON, J. 1988. Recalling the Medical Officer of Health. *Health Promotion International*, 3, 413-419.
- ASHTON, J. & SEYMOUR, H. 1988. *The new public health: the Liverpool experience*, Open University Press Milton Keynes.
- ASNER, S., WATERS, V., SOLOMON, M., YAU, Y., RICHARDSON, S. E., GRASEMANN, H., GHARABAGHI, F. & TRAN, D. 2012. Role of respiratory viruses in pulmonary exacerbations in children with cystic fibrosis. *J Cyst Fibros*, 11, 433-9.
- BALAGUER, A. & GONZALEZ DE DIOS, J. 2012. Home versus hospital intravenous antibiotic therapy for cystic fibrosis. *Cochrane Database Syst Rev*, 3, CD001917.
- BALFOUR-LYNN, I. M. & WELCH, K. 2009. Inhaled corticosteroids for cystic fibrosis. *Cochrane Database Syst Rev*, CD001915.
- BAMBRA, C., GIBSON, M., SOWDEN, A., WRIGHT, K., WHITEHEAD, M. & PETTICREW, M. 2010. Tackling the wider social determinants of health and health inequalities: evidence from systematic reviews. *J Epidemiol Community Health*, 64, 284-91.
- BAMBRA, C., GIBSON, M., SOWDEN, A. J., WRIGHT, K., WHITEHEAD, M. & PETTICREW, M. 2009. Working for health? Evidence from systematic reviews on the effects on health and health inequalities of organisational changes to the psychosocial work environment. *Prev Med*, 48, 454-61.

- BAMBRA, C., SMITH, K. E., GARTHWAITE, K., JOYCE, K. E. & HUNTER, D. J. 2011. A labour of Sisyphus? Public policy and health inequalities research from the Black and Acheson Reports to the Marmot Review. *J Epidemiol Community Health*, 65, 399-406.
- BARKER, D. J. 1997. Fetal nutrition and cardiovascular disease in later life. *Br Med Bull*, 53, 96-108.
- BARR, B., TAYLOR-ROBINSON, D. & WHITEHEAD, M. 2012. Impact on health inequalities of rising prosperity in England 1998-2007, and implications for performance incentives: longitudinal ecological study. *BMJ*, 345, e7831.
- BARR HL, S. A., FOGARTY AW 2011. The association between socioeconomic status and gender with median age at death from cystic fibrosis in England and Wales: 1959 to 2008. *BMJ*.
- BARRETT, J., DIGGLE, P., HENDERSON, R. & TAYLOR-ROBINSON, D. 2013. Joint Modelling of Repeated Measurements and Time-to-Event Outcomes: Flexible Model Specification and Exact Likelihood Inference. *Journal of the Royal Statistical Society: Series C (currently under review)*.
- BARTLEY, M. 2004. Health inequality: An introduction to theories, concepts and methods.
- BARTLEY, M. & PLEWIS, I. 1997. Does health-selective mobility account for socioeconomic differences in health? Evidence from England and Wales, 1971 to 1991. *J Health Soc Behav*, 38, 376-86.
- BAUSSANO, I., TARDIVO, I., BELLEZZA-FONTANA, R., FORNERIS, M. P., LEZO, A., ANFOSSI, L., CASTELLO, M., ALEKSANDAR, V. & BIGNAMINI, E. 2006. Neonatal screening for cystic fibrosis does not affect time to first infection with *Pseudomonas aeruginosa*. *Pediatrics*, 118, 888-95.
- BEKER, L. T., RUSSEK-COHEN, E. & FINK, R. J. 2001. Stature as a prognostic factor in cystic fibrosis survival. *J Am Diet Assoc*, 101, 438-42.
- BELLIS, G., CAZES, M. H., PARANT, A., GAIMARD, M., TRAVERS, C., LE ROUX, E., RAVILLY, S. & RAULT, G. 2007. Cystic fibrosis mortality trends in France. *J Cyst Fibros*, 6, 179-86.
- BEN-SHLOMO, Y. & KUH, D. 2002. A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *International Journal of Epidemiology*, 31, 285.
- BERGEN, H., HAWTON, K., WATERS, K., NESS, J., COOPER, J., STEEG, S. & KAPUR, N. 2012. Premature death after self-harm: a multicentre cohort study. *Lancet*, 380, 1568-74.
- BESIER, T., BORN, A., HENRICH, G., HINZ, A., QUITTNER, A. L., GOLDBECK, L. & GROUP, T. S. 2011. Anxiety, depression, and life satisfaction in parents caring for children with cystic fibrosis. *Pediatr Pulmonol*, 46, 672-82.
- BLACK, D., MORRIS, J., SMITH, C. & TOWNSEND, P. 1980. Inequalities in health: report of a Research Working Group. *London: Department of Health and Social Security*, 19.
- BLAIR, T. 1999. Beveridge lecture. *Toynbee Hall, London*, 18.
- BLANE, D., BARTLEY, M. & SMITH, G. D. 1997. Disease aetiology and materialist explanations of socioeconomic mortality differentials. *The European Journal of Public Health*, 7, 385-391.

- BLANE, D., HARDING, S. & ROSATO, M. 1999. Does social mobility affect the size of the socioeconomic mortality differential?: evidence from the Office for National Statistics Longitudinal Study. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 162, 59-70.
- BLAXTER, M. 1990. *Health and lifestyles*, Routledge.
- BOR, D. H. & EPSTEIN, P. R. 1991. Pathogenesis of respiratory infection in the disadvantaged. *Semin Respir Infect*, 6, 194-203.
- BOSWORTH, D. G. & NIELSON, D. W. 1997. Effectiveness of home versus hospital care in the routine treatment of cystic fibrosis. *Pediatr Pulmonol*, 24, 42-7.
- BOURDIEU, P. 1984. *Distinction: A social critique of the judgment of taste*, Harvard University Press.
- BRADSHAW, J. 2011. *The Well-being of Children in the UK*, The Policy Press.
- BRITTON, J. R. 1989. Effects of social class, sex, and region of residence on age at death from cystic fibrosis. *BMJ*, 298, 483-7.
- BROWNE, J. 2012. The impact of austerity measures on households with children. *Report commissioned by the Institute for Fiscal Studies*.
- BRUNNER, E. & MARMOT, M. 2006. Social organization, stress, and health. *Social determinants of health*, 2, 6-30.
- BURKER, E. J., SEDWAY, J. & CARONE, S. 2004. Psychological and educational factors: better predictors of work status than FEV1 in adults with cystic fibrosis. *Pediatr Pulmonol*, 38, 413-8.
- BUZZETTI, R., SALVATORE, D., BALDO, E., FORNERIS, M. P., LUCIDI, V., MANUNZA, D., MARINELLI, I., MESSORE, B., NERI, A. S., RAI, V., FURNARI, M. L. & MASTELLA, G. 2009. An overview of international literature from cystic fibrosis registries: 1. Mortality and survival studies in cystic fibrosis. *J Cyst Fibros*, 8, 229-37.
- CAPEWELL, S. & GRAHAM, H. 2010. Will cardiovascular disease prevention widen health inequalities? *PLoS Med*, 7, e1000320.
- CARLSEN, K., DALTON, S. O., DIDERICHSEN, F., JOHANSEN, C. & DANISH COHORT, S. 2008a. Risk for unemployment of cancer survivors: A Danish cohort study. *Eur J Cancer*, 44, 1866-74.
- CARLSEN, K., DALTON, S. O., FREDERIKSEN, K., DIDERICHSEN, F. & JOHANSEN, C. 2007. Are cancer survivors at an increased risk for divorce? A Danish cohort study. *Eur J Cancer*, 43, 2093-9.
- CARLSEN, K., OKSBJERG, D. S., FREDERIKSEN, K., DIDERICHSEN, F. & JOHANSEN, C. 2008b. Cancer and the risk for taking early retirement pension: a Danish cohort study. *Scand J Public Health*, 36, 117-25.
- CF TRUST 2008. Centre care/shared care/network care: Care of patients with cystic fibrosis in the UK. [http://www.cftrust.org.uk/aboutcf/cfcare/Model\\_of\\_CFCare.pdf](http://www.cftrust.org.uk/aboutcf/cfcare/Model_of_CFCare.pdf) (accessed 9th March 2013).
- CF TRUST 2011. Standards for the Clinical Care of Children and Adults with Cystic Fibrosis in the UK. Second edition. December 2011. [http://www.cftrust.org.uk/aboutcf/publications/consensusdoc/CF\\_Trust\\_Standards\\_of\\_Care\\_2011\\_website\\_Apr\\_12.pdf](http://www.cftrust.org.uk/aboutcf/publications/consensusdoc/CF_Trust_Standards_of_Care_2011_website_Apr_12.pdf) (accessed 7th March 2013).

- CF TRUST 2013a. CF Registry - Annual Data Reports. <https://http://www.cysticfibrosis.org.uk/about-cf/publications/cf-registry-reports.aspx> (accessed 25th April 2013).
- CF TRUST 2013b. UK Cystic Fibrosis Registry Review 2012 [https://http://www.cysticfibrosis.org.uk/media/179257/CR\\_Annual\\_Data\\_Review\\_2012\\_Apr\\_13.pdf](https://http://www.cysticfibrosis.org.uk/media/179257/CR_Annual_Data_Review_2012_Apr_13.pdf) (accessed 7th May 2013).
- CFF 2007. Cystic Fibrosis Foundation Patient Registry. Annual Data Report to the Center Directors. Bethesda, MD: Cystic Fibrosis Foundation; 2007.
- CHAMNAN, P., SHINE, B. S., HAWORTH, C. S., BILTON, D. & ADLER, A. I. 2010. Diabetes as a determinant of mortality in cystic fibrosis. *Diabetes Care*, 33, 311-6.
- CHAN, D. K., CORDATO, D., O'ROURKE, F., CHAN, D. L., POLLACK, M., MIDDLETON, S. & LEVI, C. 2013. Comprehensive stroke units: a review of comparative evidence and experience. *Int J Stroke*, 8, 260-4.
- CHANDOLA, T., BARTLEY, M., SACKER, A., JENKINSON, C. & MARMOT, M. 2003. Health selection in the Whitehall II study, UK. *Soc Sci Med*, 56, 2059-72.
- CHANDRA, A. & VOGL, T. S. 2010. Rising up with shoe leather? A comment on Fair Society, Healthy Lives (the Marmot Review). *Soc Sci Med*, 71, 1227-30; discussion 1254-8.
- CHOTIRMALL, S. H., SMITH, S. G., GUNARATNAM, C., COSGROVE, S., DIMITROV, B. D., O'NEILL, S. J., HARVEY, B. J., GREENE, C. M. & MCELVANEY, N. G. 2012. Effect of estrogen on pseudomonas mucoidy and exacerbations in cystic fibrosis. *N Engl J Med*, 366, 1978-86.
- CLAUSSEN, B., SMITS, J., NAESS, O. & DAVEY SMITH, G. 2005. Intragenerational mobility and mortality in Oslo: social selection versus social causation. *Soc Sci Med*, 61, 2513-20.
- COHEN-CYMBERKNOH, M., SHOSEYOV, D. & KEREM, E. 2011. Managing cystic fibrosis: strategies that increase life expectancy and improve quality of life. *Am J Respir Crit Care Med*, 183, 1463-71.
- COLLACO, J. M., BLACKMAN, S. M., MCGREADY, J., NAUGHTON, K. M. & CUTTING, G. R. 2010. Quantification of the relative contribution of environmental and genetic factors to variation in cystic fibrosis lung function. *J Pediatr*, 157, 802-7 e1-3.
- COLLACO, J. M., MCGREADY, J., GREEN, D. M., NAUGHTON, K. M., WATSON, C. P., SHIELDS, T., BELL, S. C., WAINWRIGHT, C. E., GROUP, A. S. & CUTTING, G. R. 2011. Effect of temperature on cystic fibrosis lung disease and infections: a replicated cohort study. *PLoS One*, 6, e27784.
- COLLACO, J. M., VANSKOY, L., BREMER, L., MCDUGAL, K., BLACKMAN, S. M., BOWERS, A., NAUGHTON, K., JENNINGS, J., ELLEN, J. & CUTTING, G. R. 2008. Interactions between secondhand smoke and genes that affect cystic fibrosis lung disease. *JAMA*, 299, 417-24.
- COOPER, H., SMAJE, C. & ARBER, S. 1998. Use of health services by children and young people according to ethnicity and social class: secondary analysis of a national survey. *BMJ*, 317, 1047-51.
- COOPER, P. J., ROBERTSON, C. F., HUDSON, I. L. & PHELAN, P. D. 1990. Variability of pulmonary function tests in cystic fibrosis. *Pediatr Pulmonol*, 8, 16-22.
- COREY, M. 1996. Survival estimates in cystic fibrosis: snapshots of a moving target. *Pediatr Pulmonol*, 21, 149-50.

- COREY, M. 2007. Power considerations for studies of lung function in cystic fibrosis. *Proc Am Thorac Soc*, 4, 334-7.
- COREY, M., EDWARDS, L., LEVISON, H. & KNOWLES, M. 1997. Longitudinal analysis of pulmonary function decline in patients with cystic fibrosis. *J Pediatr*, 131, 809-14.
- COREY, M. & FAREWELL, V. 1996. Determinants of mortality from cystic fibrosis in Canada, 1970-1989. *Am J Epidemiol*, 143, 1007-17.
- COREY, M., MCLAUGHLIN, F. J., WILLIAMS, M. & LEVISON, H. 1988. A comparison of survival, growth, and pulmonary function in patients with cystic fibrosis in Boston and Toronto. *J Clin Epidemiol*, 41, 583-91.
- CROSS-GOVERNMENT OBESITY UNIT 2008. Department of Health, Department of Children, Schools and Families. Healthy Weight, Healthy Lives: A Cross- Government Strategy for England. London: Cross-Government Obesity Unit 2008. [http://webarchive.nationalarchives.gov.uk/20100407220245/http://www.dh.gov.uk/prod\\_consum\\_dh/groups/dh\\_digitalassets/documents/digitalasset/dh\\_084024.pdf](http://webarchive.nationalarchives.gov.uk/20100407220245/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_084024.pdf) (accessed 5th March 2013).
- CRUZ, I., MARCIEL, K. K., QUITTNER, A. L. & SCHECHTER, M. S. 2009. Anxiety and depression in cystic fibrosis. *Semin Respir Crit Care Med*, 30, 569-78.
- CSDH 2008. Closing the gap in a generation: health equity through action on the social determinants of health. Final Report of the Commission on Social Determinants of Health. [http://whqlibdoc.who.int/publications/2008/9789241563703\\_eng.pdf](http://whqlibdoc.who.int/publications/2008/9789241563703_eng.pdf) (last accessed 29th August 2008).
- CUTTING, G. R. 2010. Modifier genes in Mendelian disorders: the example of cystic fibrosis. *Ann N Y Acad Sci*, 1214, 57-69.
- CYSTIC FIBROSIS TRUST 2011. UK CF Registry annual data report 2010. <http://www.cftrust.org.uk/aboutcf/publications/cfregistryreports/> (accessed 28th September 2012).
- DAHL, E. 1993. Social inequality in health--the role of the healthy worker effect. *Soc Sci Med*, 36, 1077-86.
- DAHLGREN, G. & WHITEHEAD, M. Tackling inequalities in health: what can we learn from what has been tried. Technical input paper for King's Fund International Seminar on Tackling Inequalities in Health. . 1993 Oxford: Ditchley Park.
- DAHLGREN, G. & WHITEHEAD, M. 2007. European strategies for tackling social inequities in health: Levelling up Part 2. *Copenhagen: World Health Organization*.
- DARZI, A. R. 2008. High quality care for all: NHS Next Stage Review final report. *London, Department of Health*.
- DAVENPORT, C. B. 1916. The hereditary factor in pellagra. *Archives of Internal Medicine*, 18, 4.
- DAVIES, J. C., ALTON, E. & BUSH, A. 2007. Clinical Review: Cystic fibrosis. *BMJ: British Medical Journal*, 335, 1255.
- DAVIES, J. C. & ALTON, E. W. 2009. Monitoring respiratory disease severity in cystic fibrosis. *Respir Care*, 54, 606-17.
- DE JONG, P. A., LINDBLAD, A., RUBIN, L., HOP, W. C., DE JONGSTE, J. C., BRINK, M. & TIDDENS, H. A. 2006. Progression of lung disease on computed

- tomography and pulmonary function tests in children and adults with cystic fibrosis. *Thorax*, 61, 80-5.
- DEMKO, C. A., BYARD, P. J. & DAVIS, P. B. 1995. Gender differences in cystic fibrosis: Pseudomonas aeruginosa infection. *J Clin Epidemiol*, 48, 1041-9.
- DEPARTMENT FOR EDUCATION AND SKILLS 2004. Every Child Matters. Nottingham: Department for Education and Skills. <https://http://www.education.gov.uk/publications/standard/publicationDetail/Page1/DfES/1081/2004> (accessed 5th March 2013).
- DEPARTMENT OF WORK AND PENSIONS 2011. Disability Living Allowance reform - Submission by the Cystic Fibrosis Trust <http://www.dwp.gov.uk/docs/dla-reform-cystic-fibrosis-trust.doc> (accessed 25th April 2013).
- DEX, S. & JOSHI, H. 2004. *Millennium Cohort Study First Survey: a user's guide to initial findings*, Centre for Longitudinal Studies, Institute of Education, University of London.
- DH 1998a. Independent inquiry into inequalities in health (the Acheson Report) [http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH\\_4097582](http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4097582) (accessed 5th March 2013).
- DH 1998b. Smoking Kills: A White Paper on Tobacco. London: The Stationery Office 1998. [http://webarchive.nationalarchives.gov.uk/+www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH\\_4006684](http://webarchive.nationalarchives.gov.uk/+www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4006684) (accessed 5th March 2013).
- DH 1999. Reducing Health Inequalities: An Action Report. London: Department of Health, 1999. [http://webarchive.nationalarchives.gov.uk/+www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH\\_4006054](http://webarchive.nationalarchives.gov.uk/+www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4006054) (accessed 5th March 2013).
- DH 2003. Tackling Health Inequalities: A Program for Action. London: Department of Health, 2003. [http://webarchive.nationalarchives.gov.uk/+www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH\\_4008268](http://webarchive.nationalarchives.gov.uk/+www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_4008268) (accessed 5th March 2013).
- DH 2012a. Health Profile, Liverpool <http://www.apho.org.uk/resource/item.aspx?RID=117032> (accessed 19th Feb 2013).
- DH 2012b. Improving outcomes and supporting transparency. Part 1: A public health outcomes framework for England, 2013-2016 [http://www.dh.gov.uk/prod\\_consum\\_dh/groups/dh\\_digitalassets/@dh/@en/documents/digitalasset/dh\\_132559.pdf](http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_132559.pdf) (accessed 3 December 2012).
- DH 2012c. Reducing health inequalities – The Health and Social Care Act 2012 [https://http://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/138267/C2.-Factsheet-Tackling-inequalities-in-healthcare-270412.pdf](https://http://www.gov.uk/government/uploads/system/uploads/attachment_data/file/138267/C2.-Factsheet-Tackling-inequalities-in-healthcare-270412.pdf) (accessed 12th April 2013).
- DIDERICHSEN, F. Understanding health equity in populations—some theoretical and methodological considerations. Promoting Research on Inequality in Health. Proceedings from an international expert meeting. Stockholm: Swedish Council for Social Research, 1998. 203-18.

- DIDERICHSEN, F., EVANS, T. & WHITEHEAD, M. 2001. The social origins of disparities in health. *Challenging inequities in health*. New York: Oxford University Press.
- DIGGLE, P., HEAGERTY, P., LIANG, K.-Y. & ZEGER, S. L. 2002. *Analysis of Longitudinal Data (second edition)*, Oxford, Oxford University Press.
- DODGE, J. A., LEWIS, P. A., STANTON, M. & WILSHER, J. 2007. Cystic fibrosis mortality and survival in the UK: 1947-2003. *European Respiratory Journal*, 29, 522-526.
- DODGE, J. A., MORISON, S., LEWIS, P. A., COLES, E. C., GEDDES, D., RUSSELL, G., LITTLEWOOD, J. M. & SCOTT, M. T. 1997. Incidence, population, and survival of cystic fibrosis in the UK, 1968-95. UK Cystic Fibrosis Survey Management Committee. *Arch Dis Child*, 77, 493-6.
- DODGE, J. A., MORISON, S., LEWIS, P. A., COLEST, E. C., GEDDES, D., RUSSELL, G., JACKSON, A. D. & BENTLEY, B. 1993. Cystic fibrosis in the United Kingdom, 1968-1988: incidence, population and survival. *Paediatr Perinat Epidemiol*, 7, 157-66.
- DOULL, I., EVANS, H., SOUTH & MID WALES PAEDIATRIC CYSTIC FIBROSIS, N. 2012. Full, shared and hybrid paediatric care for cystic fibrosis in South and Mid Wales. *Arch Dis Child*, 97, 17-20.
- DRUMM, M. L., KONSTAN, M. W., SCHLUCHTER, M. D., HANDLER, A., PACE, R., ZOU, F., ZARIWALA, M., FARGO, D., XU, A., DUNN, J. M., DARRAH, R. J., DORFMAN, R., SANDFORD, A. J., COREY, M., ZIELENSKI, J., DURIE, P., GODDARD, K., YANKASKAS, J. R., WRIGHT, F. A., KNOWLES, M. R. & GENE MODIFIER STUDY, G. 2005. Genetic modifiers of lung disease in cystic fibrosis. *N Engl J Med*, 353, 1443-53.
- EDWARDS, J. & BOXALL, K. 2010. Adults with cystic fibrosis and barriers to employment. *Disability & Society*, 25, 441-453.
- EDWARDS, L. J. 2000. Modern statistical techniques for the analysis of longitudinal data in biomedical research. *Pediatr Pulmonol*, 30, 330-44.
- ELBORN, J. S., SHALE, D. J. & BRITTON, J. R. 1991. Cystic fibrosis: current survival and population estimates to the year 2000. *Thorax*, 46, 881-5.
- ELLISON, A. M. & BAUCHNER, H. 2007. Socioeconomic status and length of hospital stay in children with vaso-occlusive crises of sickle cell disease. *J Natl Med Assoc*, 99, 192-6.
- EMERSON, J., ROSENFELD, M., MCNAMARA, S., RAMSEY, B. & GIBSON, R. L. 2002. Pseudomonas aeruginosa and other predictors of mortality and morbidity in young children with cystic fibrosis. *Pediatr Pulmonol*, 34, 91-100.
- ENGELS, F. 1845. *The condition of the working class in England*, Penguin.
- ENRIGHT, P. L., BECK, K. C. & SHERRILL, D. L. 2004. Repeatability of spirometry in 18,000 adult patients. *Am J Respir Crit Care Med*, 169, 235-8.
- ERIKSSON, J. G., FORSEN, T., TUOMILEHTO, J., OSMOND, C. & BARKER, D. J. 2001. Early growth and coronary heart disease in later life: longitudinal study. *BMJ*, 322, 949-53.
- EUROPEAN MEDICINES AGENCY 2009. Committee for Medicinal Products for Human Use (CHMP). Guideline on the clinical development of medicinal products for the treatment of cystic fibrosis. EMEA/CHMP/EWP/9147/2008, October, 2009.

[http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Scientific\\_guideline/2009/12/WC500017055.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/12/WC500017055.pdf) (accessed 4th April 2013).

- FARHAT, S. C., ALMEIDA, M. B., SILVA-FILHO, L. V., FARHAT, J., RODRIGUES, J. C. & BRAGA, A. L. 2013. Ozone Is Associated with an Increased Risk of Respiratory Exacerbations in Cystic Fibrosis Patients. *Chest*.
- FERGUSON, H., BOVAIRD, S. & MUELLER, M. 2007. The impact of poverty on educational outcomes for children. *Paediatr Child Health*, 12, 701-6.
- FIELD, F. 2010. *The Foundation Years: preventing poor children becoming poor adults*, The Stationery Office/Tso.
- FITZMAURICE, G. M., LAIRD, N.M. AND WARE, J.H. 2004. *Applied Longitudinal Analysis*, New Jersey, Wiley.
- FITZMAURICE, G. M. & RAVICHANDRAN, C. 2008. A primer in longitudinal data analysis. *Circulation*, 118, 2005-10.
- FLUME, P. A., MOGAYZEL, P. J., JR., ROBINSON, K. A., GOSS, C. H., ROSENBLATT, R. L., KUHN, R. J., MARSHALL, B. C. & CLINICAL PRACTICE GUIDELINES FOR PULMONARY THERAPIES, C. 2009a. Cystic fibrosis pulmonary guidelines: treatment of pulmonary exacerbations. *Am J Respir Crit Care Med*, 180, 802-8.
- FLUME, P. A., MOGAYZEL, P. J., JR., ROBINSON, K. A., ROSENBLATT, R. L., QUITTELL, L., MARSHALL, B. C., CLINICAL PRACTICE GUIDELINES FOR PULMONARY THERAPIES, C. & CYSTIC FIBROSIS FOUNDATION PULMONARY THERAPIES, C. 2010. Cystic fibrosis pulmonary guidelines: pulmonary complications: hemoptysis and pneumothorax. *Am J Respir Crit Care Med*, 182, 298-306.
- FLUME, P. A., O'SULLIVAN, B. P., ROBINSON, K. A., GOSS, C. H., MOGAYZEL, P. J., JR., WILLEY-COURAND, D. B., BUJAN, J., FINDER, J., LESTER, M., QUITTELL, L., ROSENBLATT, R., VENDER, R. L., HAZLE, L., SABADOSA, K., MARSHALL, B. & CYSTIC FIBROSIS FOUNDATION, P. T. C. 2007. Cystic fibrosis pulmonary guidelines: chronic medications for maintenance of lung health. *Am J Respir Crit Care Med*, 176, 957-69.
- FLUME, P. A., ROBINSON, K. A., O'SULLIVAN, B. P., FINDER, J. D., VENDER, R. L., WILLEY-COURAND, D. B., WHITE, T. B., MARSHALL, B. C. & CLINICAL PRACTICE GUIDELINES FOR PULMONARY THERAPIES, C. 2009b. Cystic fibrosis pulmonary guidelines: airway clearance therapies. *Respir Care*, 54, 522-37.
- FOGARTY, A., HUBBARD, R. & BRITTON, J. 2000. International comparison of median age at death from cystic fibrosis. *Chest*, 117, 1656-60.
- FOGARTY, A. W., BRITTON, J., CLAYTON, A. & SMYTH, A. R. 2012. Are measures of body habitus associated with mortality in cystic fibrosis? *Chest*, 142, 712-7.
- FOGARTY, A. W., LEWIS, S. A., MCKEEVER, T. M. & BRITTON, J. 2011. The association of two different measures of body habitus with lung function: a population-based study. *Respir Med*, 105, 1896-901.
- FOLLERAS, S., MICHALSEN, H. & BENTSEN, B. S. 1988. Social-medical aspects of cystic fibrosis in Norway. II. Economy and housing situation. *Scand J Gastroenterol Suppl*, 143, 56-9.
- FPH 2010. FPH calls on government to withdraw Health and Social Care Bill 'in best interests of everyone's health'  
[http://www.fph.org.uk/fph\\_calls\\_on\\_government\\_to\\_withdraw\\_health\\_and\\_s](http://www.fph.org.uk/fph_calls_on_government_to_withdraw_health_and_s)

- [ocial care bill 'in best interests of everyone's health'](#) (accessed 12th March 2013).
- FRANGOLIAS, D. D., HOLLOWAY, C. L., VEDAL, S. & WILCOX, P. G. 2003. Role of exercise and lung function in predicting work status in cystic fibrosis. *Am J Respir Crit Care Med*, 167, 150-7.
- FRAZER, W. 1947. Duncan of Liverpool.
- FREDEKSEN, B., LANNG, S., KOCH, C. & HOIBY, N. 1996. Improved survival in the Danish center-treated cystic fibrosis patients: results of aggressive treatment. *Pediatr Pulmonol*, 21, 153-8.
- FUCHS, H. J., BOROWITZ, D. S., CHRISTIANSEN, D. H., MORRIS, E. M., NASH, M. L., RAMSEY, B. W., ROSENSTEIN, B. J., SMITH, A. L. & WOHL, M. E. 1994. Effect of aerosolized recombinant human DNase on exacerbations of respiratory symptoms and on pulmonary function in patients with cystic fibrosis. The Pulmozyme Study Group. *N Engl J Med*, 331, 637-42.
- GALOBARDES, B. 2012. Socioeconomic inequalities in health: individual or area level; does it matter? *BMC Public Health*, 12, 171.
- GALOBARDES, B., LYNCH, J. & SMITH, G. D. 2007. Measuring socioeconomic position in health research. *Br Med Bull*, 81-82, 21-37.
- GALOBARDES, B., LYNCH, J. W. & DAVEY SMITH, G. 2004. Childhood socioeconomic circumstances and cause-specific mortality in adulthood: systematic review and interpretation. *Epidemiol Rev*, 26, 7-21.
- GALOBARDES, B., LYNCH, J. W. & SMITH, G. D. 2008. Is the association between childhood socioeconomic circumstances and cause-specific mortality established? Update of a systematic review. *J Epidemiol Community Health*, 62, 387-90.
- GALOBARDES, B., SHAW, M., LAWLOR, D. A., LYNCH, J. W. & DAVEY SMITH, G. 2006a. Indicators of socioeconomic position (part 1). *J Epidemiol Community Health*, 60, 7-12.
- GALOBARDES, B., SHAW, M., LAWLOR, D. A., LYNCH, J. W. & DAVEY SMITH, G. 2006b. Indicators of socioeconomic position (part 2). *J Epidemiol Community Health*, 60, 95-101.
- GEDDES, I., ALLEN, J., ALLEN, M. & MORRISEY, L. 2011. The Marmot Review: implications for spatial planning. *London: The Marmot Review Team*.
- GELLER, D. E. & MADGE, S. 2011. Technological and behavioral strategies to reduce treatment burden and improve adherence to inhaled antibiotics in cystic fibrosis. *Respir Med*, 105 Suppl 2, S24-31.
- GEOCONVERT. 2011. <http://geoconvert.mimas.ac.uk/> (accessed 29th July 2011) [Online].
- GEORGE, P. M., BANYA, W., PAREEK, N., BILTON, D., CULLINAN, P., HODSON, M. E. & SIMMONDS, N. J. 2011. Improved survival at low lung function in cystic fibrosis: cohort study from 1990 to 2007. *BMJ*, 342, d1008.
- GERONIMUS, A. T. 2006. Invited commentary: Using area-based socioeconomic measures--think conceptually, act cautiously. *Am J Epidemiol*, 164, 835-40; discussion 841-3.
- GIBBS, J. L. & CUNNINGHAM, D. 2002. Mortality and volume of cases in paediatric cardiac surgery. Volunteered mortality data may be unreliable. *BMJ*, 324, 1095.
- GILLEN, M., LALLAS, D., BROWN, C., YELIN, E. & BLANC, P. 1995. Work disability in adults with cystic fibrosis. *Am J Respir Crit Care Med*, 152, 153-6.

- GLIKLICH, R. & DREYER, N. 2010. Registries for Evaluating Patient Outcomes: A User's Guide. Rockville, MD: Agency for Healthcare Research and Quality Publication; September 2010.(AHRQ Publication; 10-EHC049).
- GOEMINNE, P., KICINSKI, M., VERMEULEN, F., FIERENS, F., DE BOECK, K., NEMERY, B., NAWROT, T. S. & DUPONT, L. J. 2012. Impact of air pollution on cystic fibrosis pulmonary exacerbations: a case-crossover analysis. *Chest*, 143, 946-54.
- GOLDMAN, D. P. & SMITH, J. P. 2002. Can patient self-management help explain the SES health gradient? *Proc Natl Acad Sci U S A*, 99, 10929-34.
- GOLDMAN, N. 2001. Social inequalities in health disentangling the underlying mechanisms. *Ann N Y Acad Sci*, 954, 118-39.
- GOSS, C. H. & BURNS, J. L. 2007. Exacerbations in cystic fibrosis. 1: Epidemiology and pathogenesis. *Thorax*, 62, 360-7.
- GOSS, C. H., NEWSOM, S. A., SCHILDCROUT, J. S., SHEPPARD, L. & KAUFMAN, J. D. 2004. Effect of ambient air pollution on pulmonary exacerbations and lung function in cystic fibrosis. *Am J Respir Crit Care Med*, 169, 816-21.
- GRAHAM, H. 2004. Social determinants and their unequal distribution: clarifying policy understandings. *Milbank Quarterly*, 82, 101-124.
- GRAHAM, H. 2007. *Unequal lives*, Open University Press.
- GREASEMANN, H. & RATJEN, F. 2013. Early lung disease in cystic fibrosis. *The Lancet Respiratory Medicine*. doi:10.1016/S2213-2600(13)70026-2.
- GRIMES, D. A. & SCHULZ, K. F. 2008. Making sense of odds and odds ratios. *Obstet Gynecol*, 111, 423-6.
- GROHOLT, E. K., STIGUM, H., NORDHAGEN, R. & KOHLER, L. 2003. Health service utilization in the Nordic countries in 1996: Influence of socio-economic factors among children with and without chronic health conditions. *Eur J Public Health*, 13, 30-7.
- GUPTA, A., URQUHART, D. & ROSENTHAL, M. 2009. Marked improvement in cystic fibrosis lung disease and nutrition following change in home environment. *J R Soc Med*, 102 Suppl 1, 45-8.
- HALFON, N. & HOCHSTEIN, M. 2002. Life course health development: an integrated framework for developing health, policy, and research. *Milbank Q*, 80, 433-79, iii.
- HALFON, N., INKELAS, M. & HOCHSTEIN, M. 2000. The health development organization: an organizational approach to achieving child health development. *Milbank Q*, 78, 447-97, 341.
- HALLDORSSON, M., KUNST, A. E., KOHLER, L. & MACKENBACH, J. P. 2002. Socioeconomic differences in children's use of physician services in the Nordic countries. *J Epidemiol Community Health*, 56, 200-4.
- HALONEN, J. I., KIVIMAKI, M., PENTTI, J., KAWACHI, I., VIRTANEN, M., MARTIKAINEN, P., SUBRAMANIAN, S. V. & VAHTERA, J. 2012. Quantifying neighbourhood socioeconomic effects in clustering of behaviour-related risk factors: a multilevel analysis. *PLoS One*, 7, e32937.
- HANKINSON, J. L., ODENCRANTZ, J. R. & FEDAN, K. B. 1999. Spirometric reference values from a sample of the general U.S. population. *Am J Respir Crit Care Med*, 159, 179-87.
- HANRATTY, B., BURSTROM, B., WALANDER, A. & WHITEHEAD, M. 2007a. Inequality in the face of death? Public expenditure on health care for

- different socioeconomic groups in the last year of life. *J Health Serv Res Policy*, 12, 90-4.
- HANRATTY, B., ZHANG, T. & WHITEHEAD, M. 2007b. How close have universal health systems come to achieving equity in use of curative services? A systematic review. *Int J Health Serv*, 37, 89-109.
- HARBER, P., TASHKIN, D. P., SIMMONS, M., CRAWFORD, L., HNIZDO, E. & CONNETT, J. 2007. Effect of occupational exposures on decline of lung function in early chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*, 176, 994-1000.
- HART, J. T. 1971. The inverse care law. *Lancet*, 1, 405-12.
- HASTIE, T. J. & TIBSHIRANI, R. J. 1990. *Generalized additive models*, Chapman & Hall/CRC.
- HAVERMANS, T., COLPAERT, K., VANHAREN, L. & DUPONT, L. J. 2009. Health related quality of life in cystic fibrosis: To work or not to work? *J Cyst Fibros*, 8, 218-23.
- HERRNSTEIN, R. J. 1971. IQ Atlantic Monthly. *September (228: 7)*, 43 - 64.
- HIATT, P. W., GRACE, S. C., KOZINETZ, C. A., RABOUDI, S. H., TREECE, D. G., TABER, L. H. & PIEDRA, P. A. 1999. Effects of viral lower respiratory tract infection on lung function in infants with cystic fibrosis. *Pediatrics*, 103, 619-26.
- HIJMANS, C. T., FIJNVANDRAAT, K., OOSTERLAAN, J., HEIJBOER, H., PETERS, M. & GROOTENHUIS, M. A. 2010. Double disadvantage: a case control study on health-related quality of life in children with sickle cell disease. *Health Qual Life Outcomes*, 8, 121.
- HILL, A. B. 1965. The Environment and Disease: Association or Causation? *Proc R Soc Med*, 58, 295-300.
- HIMSWORTH, H. 1984. Epidemiology, genetics and sociology. *J Biosoc Sci*, 16, 159-76.
- HIRSCH, D. 2008. Estimating the costs of child poverty. *The Joseph Rowntree Foundation. York*.
- HIRSCH D, S. N. 2008. Health Consequences of Poverty for Children. [http://www.endchildpoverty.org.uk/files/Health\\_consequences\\_of\\_Poverty\\_for\\_children.pdf](http://www.endchildpoverty.org.uk/files/Health_consequences_of_Poverty_for_children.pdf).
- HNIZDO, E., YU, L., FREYDER, L., ATTFIELD, M., LEFANTE, J. & GLINDMEYER, H. W. 2005. The precision of longitudinal lung function measurements: monitoring and interpretation. *Occup Environ Med*, 62, 695-701.
- HOFFMAN, C. & PARADISE, J. 2008. Health insurance and access to health care in the United States. *Ann N Y Acad Sci*, 1136, 149-60.
- HOGG, M., BRAITHWAITE, M., BAILEY, M., KOTSIMBOS, T. & WILSON, J. W. 2007. Work disability in adults with cystic fibrosis and its relationship to quality of life. *J Cyst Fibros*, 6, 223-7.
- HOLGATE, S. T. 2007. Priorities for respiratory research in the UK. *Thorax*, 62, 5-7.
- HOLLAND, P., BURSTROM, B., MOLLER, I. & WHITEHEAD, M. 2006. Gender and socio-economic variations in employment among patients with a diagnosed musculoskeletal disorder: a longitudinal record linkage study in Sweden. *Rheumatology (Oxford)*, 45, 1016-22.
- HOLLAND, P., BURSTROM, B., MOLLER, I. & WHITEHEAD, M. 2009. Socioeconomic inequalities in the employment impact of ischaemic heart

- disease: a longitudinal record linkage study in Sweden. *Scand J Public Health*, 37, 450-8.
- HOLLAND, P., BURSTROM, B., WHITEHEAD, M., DIDERICHSEN, F., DAHL, E., BARR, B., NYLEN, L., CHEN, W. H., THIELEN, K., VAN DER WEL, K. A., CLAYTON, S. & UPPAL, S. 2011a. How do macro-level contexts and policies affect the employment chances of chronically ill and disabled people? Part I: The impact of recession and deindustrialization. *Int J Health Serv*, 41, 395-413.
- HOLLAND, P., NYLEN, L., THIELEN, K., VAN DER WEL, K. A., CHEN, W. H., BARR, B., BURSTROM, B., DIDERICHSEN, F., ANDERSEN, P. K., DAHL, E., UPPAL, S., CLAYTON, S. & WHITEHEAD, M. 2011b. How do macro-level contexts and policies affect the employment chances of chronically ill and disabled people? Part II: The impact of active and passive labor market policies. *Int J Health Serv*, 41, 415-30.
- HOLTZMAN, N. A. 2002. Genetics and social class. *J Epidemiol Community Health*, 56, 529-35.
- HOWE, L. D., TILLING, K., GALO BARDES, B., SMITH, G. D., GUNNELL, D. & LAWLOR, D. A. 2012. Socioeconomic differences in childhood growth trajectories: at what age do height inequalities emerge? *J Epidemiol Community Health*, 66, 143-8.
- HOWE, L. D., TILLING, K., GALO BARDES, B., SMITH, G. D., NESS, A. R. & LAWLOR, D. A. 2011. Socioeconomic disparities in trajectories of adiposity across childhood. *Int J Pediatr Obes*, 6, e144-53.
- HRUBY, J. & BUTLER, J. 1975. Variability of routine pulmonary function tests. *Thorax*, 30, 548-53.
- INSTITUTE OF MEDICINE 2013. U.S. Health in International Perspective: Shorter Lives, Poorer Health <http://www.iom.edu/Reports/2013/US-Health-in-International-Perspective-Shorter-Lives-Poorer-Health.aspx> (accessed 5th March 2013).
- JACOBS, P. & MCDERMOTT, S. 1989. Family caregiver costs of chronically ill and handicapped children: method and literature review. *Public Health Rep*, 104, 158-63.
- JADIN, S. A., WU, G. S., ZHANG, Z., SHOFF, S. M., TIPPETS, B. M., FARRELL, P. M., MILLER, T., ROCK, M. J., LEVY, H. & LAI, H. J. 2011. Growth and pulmonary outcomes during the first 2 y of life of breastfed and formula-fed infants diagnosed with cystic fibrosis through the Wisconsin Routine Newborn Screening Program. *Am J Clin Nutr*, 93, 1038-47.
- JAFFE, S. 2012. US Supreme Court makes historic health ruling. *Lancet*, 380, 14.
- JASSAL, M. S., YU, A. M., BHATIA, R., KEENS, T. G. & DAVIDSON WARD, S. L. 2013. Effect of residential proximity to major roadways on cystic fibrosis exacerbations. *Int J Environ Health Res*, 23, 119-31.
- JEFFERIS, B. J., POWER, C., GRAHAM, H. & MANOR, O. 2004. Effects of childhood socioeconomic circumstances on persistent smoking. *Am J Public Health*, 94, 279-85.
- JOHNSON, C., BUTLER, S. M., KONSTAN, M. W., MORGAN, W. & WOHL, M. E. 2003. Factors influencing outcomes in cystic fibrosis: a center-based analysis. *Chest*, 123, 20-7.
- JONES, A. M., DODD, M. E., GOVAN, J. R., DOHERTY, C. J., SMITH, C. M., ISALSKA, B. J. & WEBB, A. K. 2005. Prospective surveillance for *Pseudomonas*

- aeruginosa cross-infection at a cystic fibrosis center. *Am J Respir Crit Care Med*, 171, 257-60.
- JONES, A. P. & WALLIS, C. 2010. Dornase alfa for cystic fibrosis. *Cochrane Database Syst Rev*, CD001127.
- JUDGE, K. 1995. Income distribution and life expectancy: a critical appraisal. *BMJ*, 311, 1282-5; discussion 1285-7.
- JUDGE, K., PLATT, S., COSTONGS, C. & JURCZAK, K. 2005. *Health inequalities: a challenge for Europe*, Produced by COI for the UK Presidency of the EU.
- JUDGE, L. 2012. Ending Child Poverty by 2020: Progress made and lessons learned.
- KALER, S. G. 2008. Diseases of poverty with high mortality in infants and children: malaria, measles, lower respiratory infections, and diarrheal illnesses. *Ann N Y Acad Sci*, 1136, 28-31.
- KASTNER-COLE, D., PALMER, C. N., OGSTON, S. A., MEHTA, A. & MUKHOPADHYAY, S. 2005. Overweight and obesity in deltaF508 homozygous cystic fibrosis. *J Pediatr*, 147, 402-4.
- KAWACHI, I., SUBRAMANIAN, S. V. & ALMEIDA-FILHO, N. 2002. A glossary for health inequalities. *J Epidemiol Community Health*, 56, 647-52.
- KELLY, Y. J. & WATT, R. G. 2005. Breast-feeding initiation and exclusive duration at 6 months by social class--results from the Millennium Cohort Study. *Public Health Nutr*, 8, 417-21.
- KEREM, B., ROMMENS, J. M., BUCHANAN, J. A., MARKIEWICZ, D., COX, T. K., CHAKRAVARTI, A., BUCHWALD, M. & TSUI, L. C. 1989. Identification of the cystic fibrosis gene: genetic analysis. *Science*, 245, 1073-80.
- KEREM, E., CONWAY, S., ELBORN, S., HEIJERMAN, H. & CONSENSUS, C. 2005. Standards of care for patients with cystic fibrosis: a European consensus. *J Cyst Fibros*, 4, 7-26.
- KEREM, E., COREY, M., KEREM, B. S., ROMMENS, J., MARKIEWICZ, D., LEVISON, H., TSUI, L. C. & DURIE, P. 1990a. The relation between genotype and phenotype in cystic fibrosis--analysis of the most common mutation (delta F508). *N Engl J Med*, 323, 1517-22.
- KEREM, E., COREY, M., STEIN, R., GOLD, R. & LEVISON, H. 1990b. Risk factors for *Pseudomonas aeruginosa* colonization in cystic fibrosis patients. *Pediatr Infect Dis J*, 9, 494-8.
- KEREM, E., REISMAN, J., COREY, M., CANNY, G. J. & LEVISON, H. 1992. Prediction of mortality in patients with cystic fibrosis. *New England Journal of Medicine*, 326, 1187-91.
- KINRA, S., BAUMER, J. H. & DAVEY SMITH, G. 2005. Early growth and childhood obesity: a historical cohort study. *Arch Dis Child*, 90, 1122-7.
- KIRKWOOD, T., BOND, J., MAY, C., MCKEITH, I. & TEH, M. 2008. Foresight mental capital and Wellbeing Project. Mental capital through life: future challenges. *Government Office for Science: London*.
- KLASEN, S. 2004. In search of the holy grail: How to achieve pro-poor growth. *Toward pro poor policies-aid, institutions, and globalization*, 63-94.
- KNOPS, R. R., VAN DALEN, E. C., MULDER, R. L., LECLERCQ, E., KNIJNENBURG, S. L., KASPERS, G. J., PIETERS, R., CARON, H. N. & KREMER, L. C. 2013. The volume effect in paediatric oncology: a systematic review. *Ann Oncol*.
- KOCH, C., CUPPENS, H., RAINISIO, M., MADESSANI, U., HARMS, H., HODSON, M., MASTELLA, G., NAVARRO, J., STRANDVIK, B., MCKENZIE, S. &

- INVESTIGATORS OF THE, E. 2001. European Epidemiologic Registry of Cystic Fibrosis (ERCF): comparison of major disease manifestations between patients with different classes of mutations. *Pediatr Pulmonol*, 31, 1-12.
- KOHANSAL, R., MARTINEZ-CAMBLOR, P., AGUSTI, A., BUIST, A. S., MANNINO, D. M. & SORIANO, J. B. 2009. The natural history of chronic airflow obstruction revisited: an analysis of the Framingham offspring cohort. *Am J Respir Crit Care Med*, 180, 3-10.
- KONDO, N., SEMBAJWE, G., KAWACHI, I., VAN DAM, R. M., SUBRAMANIAN, S. V. & YAMAGATA, Z. 2009. Income inequality, mortality, and self rated health: meta-analysis of multilevel studies. *BMJ*, 339, b4471.
- KONSTAN, M. W., BUTLER, S. M., SCHIDLOW, D. V., MORGAN, W. J., JULIUS, J. R. & JOHNSON, C. A. 1999a. Patterns of medical practice in cystic fibrosis: part I. Evaluation and monitoring of health status of patients. Investigators and Coordinators of the Epidemiologic Study of Cystic Fibrosis. *Pediatr Pulmonol*, 28, 242-7.
- KONSTAN, M. W., BUTLER, S. M., SCHIDLOW, D. V., MORGAN, W. J., JULIUS, J. R. & JOHNSON, C. A. 1999b. Patterns of medical practice in cystic fibrosis: part II. Use of therapies. Investigators and Coordinators of the Epidemiologic Study of Cystic Fibrosis. *Pediatr Pulmonol*, 28, 248-54.
- KONSTAN, M. W., BUTLER, S. M., WOHL, M. E., STODDARD, M., MATOUSEK, R., WAGENER, J. S., JOHNSON, C. A. & MORGAN, W. J. 2003. Growth and nutritional indexes in early life predict pulmonary function in cystic fibrosis. *J Pediatr*, 142, 624-30.
- KONSTAN, M. W., MORGAN, W. J., BUTLER, S. M., PASTA, D. J., CRAIB, M. L., SILVA, S. J., STOKES, D. C., WOHL, M. E., WAGENER, J. S., REGELMANN, W. E. & JOHNSON, C. A. 2007a. Risk factors for rate of decline in forced expiratory volume in one second in children and adolescents with cystic fibrosis. *J Pediatr*, 151, 134-9, 139 e1.
- KONSTAN, M. W., SCHLUCHTER, M. D., XUE, W. & DAVIS, P. B. 2007b. Clinical use of Ibuprofen is associated with slower FEV1 decline in children with cystic fibrosis. *Am J Respir Crit Care Med*, 176, 1084-9.
- KONSTAN, M. W., WAGENER, J. S. & VANDEVANTER, D. R. 2009. Characterizing aggressiveness and predicting future progression of CF lung disease. *J Cyst Fibros*, 8 Suppl 1, S15-9.
- KONSTAN, M. W., WAGENER, J. S., VANDEVANTER, D. R., PASTA, D. J., YEGIN, A., RASOULIYAN, L. & MORGAN, W. J. 2012. Risk factors for rate of decline in FEV1 in adults with cystic fibrosis. *J Cyst Fibros*, 11, 405-11.
- KOSOROK, M. R., JALALUDDIN, M., FARRELL, P. M., SHEN, G., COLBY, C. E., LAXOVA, A., ROCK, M. J. & SPLAINGARD, M. 1998. Comprehensive analysis of risk factors for acquisition of *Pseudomonas aeruginosa* in young children with cystic fibrosis. *Pediatr Pulmonol*, 26, 81-8.
- KOSOROK, M. R., WEI, W. H. & FARRELL, P. M. 1996. The incidence of cystic fibrosis. *Stat Med*, 15, 449-62.
- KOSOROK, M. R., ZENG, L., WEST, S. E., ROCK, M. J., SPLAINGARD, M. L., LAXOVA, A., GREEN, C. G., COLLINS, J. & FARRELL, P. M. 2001. Acceleration of lung disease in children with cystic fibrosis after *Pseudomonas aeruginosa* acquisition. *Pediatr Pulmonol*, 32, 277-87.

- KOVESI, T., COREY, M. & LEVISON, H. 1993. Passive smoking and lung function in cystic fibrosis. *Am Rev Respir Dis*, 148, 1266-71.
- KRAUTH, C., JALILVAND, N., WELTE, T. & BUSSE, R. 2003. Cystic fibrosis: cost of illness and considerations for the economic evaluation of potential therapies. *Pharmacoeconomics*, 21, 1001-24.
- KRIEGER, J. & HIGGINS, D. L. 2002. Housing and health: time again for public health action. *Am J Public Health*, 92, 758-68.
- KRIEGER, N., WATERMAN, P., CHEN, J. T., SOOBADER, M. J., SUBRAMANIAN, S. V. & CARSON, R. 2002. Zip code caveat: bias due to spatiotemporal mismatches between zip codes and US census-defined geographic areas--the Public Health Disparities Geocoding Project. *Am J Public Health*, 92, 1100-2.
- KUH, D., POWER, C., BLANE, D. & BARTLEY, M. 2004. A life course approach to chronic disease epidemiology. In: KUH, D., POWER, C., BLANE, D. & BARTLEY, M. (eds.) *A life course approach to chronic disease epidemiology*. Oxford University Press.
- LAI, H. C., COREY, M., FITZSIMMONS, S., KOSOROK, M. R. & FARRELL, P. M. 1999. Comparison of growth status of patients with cystic fibrosis between the United States and Canada. *Am J Clin Nutr*, 69, 531-8.
- LAI, H. J., SHOFF, S. M., FARRELL, P. M. & WISCONSIN CYSTIC FIBROSIS NEONATAL SCREENING, G. 2009. Recovery of birth weight z score within 2 years of diagnosis is positively associated with pulmonary status at 6 years of age in children with cystic fibrosis. *Pediatrics*, 123, 714-22.
- LAIRD, N. M. & WARE, J. H. 1982. Random-effects models for longitudinal data. *Biometrics*, 38, 963-74.
- LALONDE, M. 1974. A new perspective on the Health of Canadians. A working document. Ottawa: Government of Canada, 1974. ISBN 0-662-50019-9.
- LANCET 1843. Editorial, 5 August: 661.
- LANDS, L. C. & STANOJEVIC, S. 2007. Oral non-steroidal anti-inflammatory drug therapy for cystic fibrosis. *Cochrane Database Syst Rev*, CD001505.
- LANGTON HEWER, S. C. & SMYTH, A. R. 2009. Antibiotic strategies for eradicating *Pseudomonas aeruginosa* in people with cystic fibrosis. *Cochrane Database Syst Rev*, CD004197.
- LAW, C., PARKIN, C. & LEWIS, H. 2012. Policies to tackle inequalities in child health: why haven't they worked (better)? *Arch Dis Child*, 97, 301-3.
- LAWLOR, D. A., BEN-SHLOMO, Y. & LEON, D. A. 2004. Pre-adult influences on cardiovascular disease. In: KUH, D., POWER, C., BLANE, D. & BARTLEY, M. (eds.) *A Life Course Approach to Chronic Diseases Epidemiology*.
- LEDSON, M. J., GALLAGHER, M. J., JACKSON, M., HART, C. A. & WALSHAW, M. J. 2002. Outcome of *Burkholderia cepacia* colonisation in an adult cystic fibrosis centre. *Thorax*, 57, 142-5.
- LEEM, J. H., KIM, H. C., LEE, J. Y. & SOHN, J. R. 2011. Interaction between bronchiolitis diagnosed before 2 years of age and socio-economic status for bronchial hyperreactivity. *Environ Health Toxicol*, 26, e2011012.
- LEVY, H., KALISH, L. A., CANNON, C. L., GARCIA, K. C., GERARD, C., GOLDMANN, D., PIER, G. B., WEISS, S. T. & COLIN, A. A. 2008. Predictors of mucoid *Pseudomonas* colonization in cystic fibrosis patients. *Pediatr Pulmonol*, 43, 463-71.

- LEVY, L., DURIE, P., PENCHARZ, P. & COREY, M. 1986. Prognostic factors associated with patient survival during nutritional rehabilitation in malnourished children and adolescents with cystic fibrosis. *J Pediatr Gastroenterol Nutr*, 5, 97-102.
- LEWIS, P. A. 1998. Inferences for health provision from survival data in cystic fibrosis. *Arch Dis Child*, 79, 297-9.
- LEWIS, P. A., MORISON, S., DODGE, J. A., GEDDES, D., COLES, E. C., RUSSELL, G., LITTLEWOOD, J. M. & SCOTT, M. T. 1999. Survival estimates for adults with cystic fibrosis born in the United Kingdom between 1947 and 1967. The UK Cystic Fibrosis Survey Management Committee. *Thorax*, 54, 420-2.
- LIANG, K.-Y. & ZEGER, S. L. 1986. Longitudinal data analysis using generalized linear models. *Biometrika*, 73, 13-22.
- LIEU, T. A., RAY, G. T., FARMER, G. & SHAY, G. F. 1999. The cost of medical care for patients with cystic fibrosis in a health maintenance organization. *Pediatrics*, 103, e72.
- LIU, T. G., ADLER, F. R., FITZSIMMONS, S. C., CAHILL, B. C., HIBBS, J. R. & MARSHALL, B. C. 2001. Predictive 5-year survivorship model of cystic fibrosis. *Am J Epidemiol*, 153, 345-52.
- LIU, T. G., ELKIN, E. P., PASTA, D. J., JACOBS, J. R., KONSTAN, M. W., MORGAN, W. J. & WAGENER, J. S. 2010. Year-to-year changes in lung function in individuals with cystic fibrosis. *J Cyst Fibros*, 9, 250-6.
- LITTLE, R. J. & RUBIN, D. B. 1987. *Statistical analysis with missing data*, Wiley New York.
- LORANT, V. & BHOPAL, R. S. 2011. Ethnicity, socio-economic status and health research: insights from and implications of Charles Tilly's theory of Durable Inequality. *J Epidemiol Community Health*, 65, 671-5.
- LOZANO, R., NAGHAVI, M., FOREMAN, K., LIM, S., SHIBUYA, K., ABOYANS, V., ABRAHAM, J., ADAIR, T., AGGARWAL, R., AHN, S. Y., ALVARADO, M., ANDERSON, H. R., ANDERSON, L. M., ANDREWS, K. G., ATKINSON, C., BADDOUR, L. M., BARKER-COLLO, S., BARTELS, D. H., BELL, M. L., BENJAMIN, E. J., BENNETT, D., BHALLA, K., BIKBOV, B., BIN ABDULHAK, A., BIRBECK, G., BLYTH, F., BOLLIGER, I., BOUFOUS, S., BUCELLO, C., BURCH, M., BURNEY, P., CARAPETIS, J., CHEN, H., CHOU, D., CHUGH, S. S., COFFENG, L. E., COLAN, S. D., COLQUHOUN, S., COLSON, K. E., CONDON, J., CONNOR, M. D., COOPER, L. T., CORRIERE, M., CORTINOVIS, M., DE VACCARO, K. C., COUSER, W., COWIE, B. C., CRIQUI, M. H., CROSS, M., DABHADKAR, K. C., DAHODWALA, N., DE LEO, D., DEGENHARDT, L., DELOSSANTOS, A., DENENBERG, J., DES JARLAIS, D. C., DHARMARATNE, S. D., DORSEY, E. R., DRISCOLL, T., DUBER, H., EBEL, B., ERWIN, P. J., ESPINDOLA, P., EZZATI, M., FEIGIN, V., FLAXMAN, A. D., FOROUZANFAR, M. H., FOWKES, F. G., FRANKLIN, R., FRANSEN, M., FREEMAN, M. K., GABRIEL, S. E., GAKIDOU, E., GASPARI, F., GILLUM, R. F., GONZALEZ-MEDINA, D., HALASA, Y. A., HARING, D., HARRISON, J. E., HAVMOELLER, R., HAY, R. J., HOEN, B., HOTEZ, P. J., HOY, D., JACOBSEN, K. H., JAMES, S. L., JASRASARIA, R., JAYARAMAN, S., JOHNS, N., KARTHIKEYAN, G., KASSEBAUM, N., KEREN, A., KHOO, J. P., KNOWLTON, L. M., KOBUSINGYE, O., KORANTENG, A., KRISHNAMURTHI, R., LIPNICK, M., LIPSHULTZ, S. E., OHNO, S. L., et al. 2012. Global and regional mortality from 235 causes of

- death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*, 380, 2095-128.
- LUNDBERG, O. 1991. Childhood living conditions, health status, and social mobility: a contribution to the health selection debate. *European Sociological Review*, 7, 149-162.
- LYNCH, J. & KAPLAN, G. 2000. Socioeconomic position. In: BERKMAN, L. & KAWACHI, I. (eds.) *Social epidemiology*: New York: Oxford University Press.
- LYNCH, J. W., SMITH, G. D., KAPLAN, G. A. & HOUSE, J. S. 2000. Income inequality and mortality: importance to health of individual income, psychosocial environment, or material conditions. *BMJ*, 320, 1200-4.
- LYRATZOPOULOS, G., BARBIERE, J. M., RACHET, B., BAUM, M., THOMPSON, M. R. & COLEMAN, M. P. 2011. Changes over time in socioeconomic inequalities in breast and rectal cancer survival in England and Wales during a 32-year period (1973-2004): the potential role of health care. *Ann Oncol*, 22, 1661-6.
- MACHIN, S. & VIGNOLES, A. 2004. Educational inequality: the widening socio-economic gap. *Fiscal Studies*, 25, 107-128.
- MACKENBACH, J. P. 2005. Genetics and health inequalities: hypotheses and controversies. *J Epidemiol Community Health*, 59, 268-73.
- MACKENBACH, J. P. 2011a. Can we reduce health inequalities? An analysis of the English strategy (1997-2010). *J Epidemiol Community Health*, 65, 568-75.
- MACKENBACH, J. P. 2011b. What would happen to health inequalities if smoking were eliminated? *BMJ*, 342, d3460.
- MACKENBACH, J. P. 2012. The persistence of health inequalities in modern welfare states: the explanation of a paradox. *Soc Sci Med*, 75, 761-9.
- MACKENBACH, J. P., KUNST, A. E., CAVELAARS, A. E., GROENHOF, F. & GEURTS, J. J. 1997. Socioeconomic inequalities in morbidity and mortality in western Europe. The EU Working Group on Socioeconomic Inequalities in Health. *Lancet*, 349, 1655-9.
- MACKENBACH, J. P., MEERDING, W. J. & KUNST, A. E. 2011. Economic costs of health inequalities in the European Union. *J Epidemiol Community Health*, 65, 412-9.
- MACKENBACH, J. P., VAN DE MHEEN, H. & STRONKS, K. 1994. A prospective cohort study investigating the explanation of socio-economic inequalities in health in The Netherlands. *Soc Sci Med*, 38, 299-308.
- MACLEAN, J. E., ATENAFU, E., KIRBY-ALLEN, M., MACLUSKY, I. B., STEPHENS, D., GASEMANN, H. & SUBBARAO, P. 2008. Longitudinal decline in lung volume in a population of children with sickle cell disease. *Am J Respir Crit Care Med*, 178, 1055-9.
- MAHADEVA, R., WEBB, K., WESTERBEEK, R. C., CARROLL, N. R., DODD, M. E., BILTON, D. & LOMAS, D. A. 1998. Clinical outcome in relation to care in centres specialising in cystic fibrosis: cross sectional study. *BMJ*, 316, 1771-5.
- MANOR, O., MATTHEWS, S. & POWER, C. 2003. Health selection: the role of inter- and intra-generational mobility on social inequalities in health. *Soc Sci Med*, 57, 2217-27.

- MARMOT, M., ALLEN, J. & GOLDBLATT, P. 2010a. A social movement, based on evidence, to reduce inequalities in health. *Social Science & Medicine*, 71, 1254-1258.
- MARMOT, M. & WILKINSON, R. G. 2001. Psychosocial and material pathways in the relation between income and health: a response to Lynch et al. *BMJ*, 322, 1233-6.
- MARMOT, M. G., ALLEN, J., GOLDBLATT, P., BOYCE, T., MCNEISH, D., GRADY, M. & GEDDES, I. 2010b. Fair society, healthy lives: Strategic review of health inequalities in England post-2010. The Marmot Review: London UK. 2010.
- MARMOT, M. G., BOSMA, H., HEMINGWAY, H., BRUNNER, E. & STANSFELD, S. 1997. Contribution of job control and other risk factors to social variations in coronary heart disease incidence. *Lancet*, 350, 235-9.
- MARMOT, M. G., ROSE, G., SHIPLEY, M. & HAMILTON, P. J. 1978. Employment grade and coronary heart disease in British civil servants. *J Epidemiol Community Health*, 32, 244-9.
- MARMOT, M. G., SHIPLEY, M. J. & ROSE, G. 1984. Inequalities in death--specific explanations of a general pattern? *Lancet*, 1, 1003-6.
- MARMOT, M. G., SMITH, G. D., STANSFELD, S., PATEL, C., NORTH, F., HEAD, J., WHITE, I., BRUNNER, E. & FEENEY, A. 1991. Health inequalities among British civil servants: the Whitehall II study. *Lancet*, 337, 1387-93.
- MASELLI, J. H., SONTAG, M. K., NORRIS, J. M., MACKENZIE, T., WAGENER, J. S. & ACCURSO, F. J. 2003. Risk factors for initial acquisition of *Pseudomonas aeruginosa* in children with cystic fibrosis identified by newborn screening. *Pediatr Pulmonol*, 35, 257-62.
- MASTELLA, G., RAINISIO, M., HARMS, H. K., HODSON, M. E., KOCH, C., NAVARRO, J., STRANDVIK, B. & MCKENZIE, S. G. 2000. Allergic bronchopulmonary aspergillosis in cystic fibrosis. A European epidemiological study. Epidemiologic Registry of Cystic Fibrosis. *Eur Respir J*, 16, 464-71.
- MATERN, B. 1960. Spatial Variation. Meddelanden fran Statens Skogsforsknings Institut, Stockholm. Band 49, number 5.
- MATTHEWS, J. N. 1993. A refinement to the analysis of serial data using summary measures. *Stat Med*, 12, 27-37.
- MCCOLLUM, A. T. 1971. Cystic fibrosis: economic impact upon the family. *Am J Public Health*, 61, 1335-40.
- MCKAY, R. T., LEMASTERS, G. K., HILBERT, T. J., LEVIN, L. S., RICE, C. H., BORTON, E. K. & LOCKEY, J. E. 2011. A long term study of pulmonary function among US refractory ceramic fibre workers. *Occupational and environmental medicine*, 68, 89.
- MCKEOWN, T. 1979. *The role of medicine. Dream, mirage or Nemesis?*, Basil Blackwell Publisher Ltd., Alfred Street, Oxford OX1 4MB.
- MEHTA, G., SIMS, E. J., CULROSS, F., MCCORMICK, J. D. & MEHTA, A. 2004. Potential benefits of the UK Cystic Fibrosis Database. *J R Soc Med*, 97 Suppl 44, 60-71.
- MERELLE, M. E., SCHOUTEN, J. P., GERRITSEN, J. & DANKERT-ROELSE, J. E. 2001. Influence of neonatal screening and centralized treatment on long-term clinical outcome and survival of CF patients. *Eur Respir J*, 18, 306-15.

- MICHALSEN, H., FOLLERAS, S. & BENTSEN, B. S. 1988. Social-medical aspects of cystic fibrosis in Norway. III. The education and occupation of mothers. *Scand J Gastroenterol Suppl*, 143, 60-4.
- MILLER, M. R., HANKINSON, J., BRUSASCO, V., BURGOS, F., CASABURI, R., COATES, A., CRAPO, R., ENRIGHT, P., VAN DER GRINTEN, C. P., GUSTAFSSON, P., JENSEN, R., JOHNSON, D. C., MACINTYRE, N., MCKAY, R., NAVAJAS, D., PEDERSEN, O. F., PELLEGRINO, R., VIEGI, G. & WANGER, J. 2005. Standardisation of spirometry. *Eur Respir J*, 26, 319-38.
- MILTON, B., HOLLAND, P. & WHITEHEAD, M. 2006. The social and economic consequences of childhood-onset Type 1 diabetes mellitus across the lifecourse: a systematic review. *Diabet Med*, 23, 821-9.
- MILTON B, M. M., TAYLOR-ROBINSON D, WHITEHEAD M. 2011. How can the equity impact of universal policies be evaluated. WHOLIS E95912. Copenhagen, World Health Organisation.
- MILTON, B., WHITEHEAD, M., HOLLAND, P. & HAMILTON, V. 2004. The social and economic consequences of childhood asthma across the lifecourse: a systematic review. *Child Care Health Dev*, 30, 711-28.
- MONTGOMERY, V., OLIVER, R., REISNER, A. & FALLAT, M. E. 2002. The effect of severe traumatic brain injury on the family. *J Trauma*, 52, 1121-4.
- MORGEN, C. S., BJORK, C., ANDERSEN, P. K., MORTENSEN, L. H. & NYBO ANDERSEN, A. M. 2008. Socioeconomic position and the risk of preterm birth--a study within the Danish National Birth Cohort. *Int J Epidemiol*, 37, 1109-20.
- MORRIS, J. N., DONKIN, A. J., WONDERLING, D., WILKINSON, P. & DOWLER, E. A. 2000. A minimum income for healthy living. *J Epidemiol Community Health*, 54, 885-9.
- MORRIS, J. N., WILKINSON, P., DANGOUR, A. D., DEEMING, C. & FLETCHER, A. 2007. Defining a minimum income for healthy living (MIHL): older age, England. *Int J Epidemiol*, 36, 1300-7.
- MOTT, L. S., PARK, J., MURRAY, C. P., GANGELL, C. L., DE KLERK, N. H., ROBINSON, P. J., ROBERTSON, C. F., RANGANATHAN, S. C., SLY, P. D., STICK, S. M. & AREST, C. F. 2012. Progression of early structural lung disease in young children with cystic fibrosis assessed using CT. *Thorax*, 67, 509-16.
- MUNCK, A., DUHAMEL, J. F., LAMIREAU, T., LE LUYER, B., LE TALLEC, C., BELLON, G., ROUSSEY, M., FOUCAUD, P., GINIES, J. L., HOUZEL, A., MARGUET, C., GUILLOT, M., DAVID, V., KAPEL, N., DYARD, F. & HENNIGES, F. 2009. Pancreatic enzyme replacement therapy for young cystic fibrosis patients. *J Cyst Fibros*, 8, 14-8.
- MURRAY, C. J., VOS, T., LOZANO, R., NAGHAVI, M., FLAXMAN, A. D., MICHAUD, C., EZZATI, M., SHIBUYA, K., SALOMON, J. A., ABDALLA, S., ABOYANS, V., ABRAHAM, J., ACKERMAN, I., AGGARWAL, R., AHN, S. Y., ALI, M. K., ALVARADO, M., ANDERSON, H. R., ANDERSON, L. M., ANDREWS, K. G., ATKINSON, C., BADDOUR, L. M., BAHALIM, A. N., BARKER-COLLO, S., BARRERO, L. H., BARTELS, D. H., BASANEZ, M. G., BAXTER, A., BELL, M. L., BENJAMIN, E. J., BENNETT, D., BERNABE, E., BHALLA, K., BHANDARI, B., BIKBOV, B., BIN ABDULHAK, A., BIRBECK, G., BLACK, J. A., BLENCOWE, H., BLORE, J. D., BLYTH, F., BOLLIGER, I., BONAVENTURE, A., BOUFOUS, S., BOURNE, R., BOUSSINESQ, M., BRAITHWAITE, T., BRAYNE, C.,

- BRIDGETT, L., BROOKER, S., BROOKS, P., BRUGHA, T. S., BRYAN-HANCOCK, C., BUCELLO, C., BUCHBINDER, R., BUCKLE, G., BUDKE, C. M., BURCH, M., BURNEY, P., BURSTEIN, R., CALABRIA, B., CAMPBELL, B., CANTER, C. E., CARABIN, H., CARAPETIS, J., CARMONA, L., CELLA, C., CHARLSON, F., CHEN, H., CHENG, A. T., CHOU, D., CHUGH, S. S., COFFENG, L. E., COLAN, S. D., COLQUHOUN, S., COLSON, K. E., CONDON, J., CONNOR, M. D., COOPER, L. T., CORRIERE, M., CORTINOVIS, M., DE VACCARO, K. C., COUSER, W., COWIE, B. C., CRIQUI, M. H., CROSS, M., DABHADKAR, K. C., DAHIYA, M., DAHODWALA, N., DAMSERE-DERRY, J., DANAEI, G., DAVIS, A., DE LEO, D., DEGENHARDT, L., DELLAVALLE, R., DELOSSANTOS, A., DENENBERG, J., DERRETT, S., DES JARLAIS, D. C., DHARMARATNE, S. D., et al. 2012. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*, 380, 2197-223.
- NAKAI, M. & KE, W. 2009. Statistical Models for Longitudinal Data Analysis. *Applied Mathematical Sciences*, 3, 1979-1989.
- NASH, E. F., KAVANAGH, D., WILLIAMS, S., BIKMALLA, S., GRAY, A. & WHITEHOUSE, J. L. 2011. Implications of the current UK welfare reforms for adults with cystic fibrosis. *Clin Med*, 11, 634.
- NAVARRO, J., RAINISIO, M., HARMS, H. K., HODSON, M. E., KOCH, C., MASTELLA, G., STRANDVIK, B. & MCKENZIE, S. G. 2001. Factors associated with poor pulmonary function: cross-sectional analysis of data from the ERCF. European Epidemiologic Registry of Cystic Fibrosis. *Eur Respir J*, 18, 298-305.
- NHS. 2011. Newborn Screening for Cystic Fibrosis in England. Diagnostic outcome data and performance against UKNSPC standards. 2009 - 2010. UK Newborn Screening Programme Centre
- NIELSEN, O. H., THOMSEN, B. L., GREEN, A., ANDERSEN, P. K., HAUGE, M. & SCHIOTZ, P. O. 1988. Cystic fibrosis in Denmark 1945 to 1985. An analysis of incidence, mortality and influence of centralized treatment on survival. *Acta Paediatr Scand*, 77, 836-41.
- NOBLE, M., MCLENNAN, D., WILKINSON, K., WHITWORTH, A., EXLEY, S., BARNES, H. & DIBBEN, C. 2008. The English indices of deprivation 2007 [http://data.gov.uk/data/resource\\_cache/80/8087b9d2-2918-47c1-bedc-439bf2b2ead1/733520.pdf](http://data.gov.uk/data/resource_cache/80/8087b9d2-2918-47c1-bedc-439bf2b2ead1/733520.pdf) (accessed 24th May 2013).
- NOBLE, M., WRIGHT, G., SMITH, G. & DIBBEN, C. 2006. Measuring multiple deprivation at the small-area level. *Environment and Planning A*, 38, 169.
- O'CONNOR, G. T., QUINTON, H. B., KAHN, R., ROBICHAUD, P., MADDOCK, J., LEVER, T., DETZER, M., BROOKS, J. G. & NORTHERN NEW ENGLAND CYSTIC FIBROSIS, C. 2002. Case-mix adjustment for evaluation of mortality in cystic fibrosis. *Pediatr Pulmonol*, 33, 99-105.
- O'CONNOR, G. T., QUINTON, H. B., KNEELAND, T., KAHN, R., LEVER, T., MADDOCK, J., ROBICHAUD, P., DETZER, M. & SWARTZ, D. R. 2003. Median household income and mortality rate in cystic fibrosis. *Pediatrics*, 111, e333-9.
- O'DONNELL, D. R., PARSLOW, R. C. & DRAPER, E. S. 2010. Deprivation, ethnicity and prematurity in infant respiratory failure in PICU in the UK. *Acta Paediatr*, 99, 1186-91.

- OKANY, C. C. & AKINYANJU, O. O. 1993. The influence of socio-economic status on the severity of sickle cell disease. *Afr J Med Med Sci*, 22, 57-60.
- ONS 2011a. Indices of Deprivation across the UK <http://www.neighbourhood.statistics.gov.uk/dissemination/Info.do?page=analysisandguidance/analysisarticles/indices-of-deprivation.htm> (accessed 29th July 2011).
- ONS 2011b. Life expectancy at birth and at age 65 by local areas in the United Kingdom, 2004-06 to 2008-10 <http://www.ons.gov.uk/ons/publications/reference-tables.html?edition=tcm%3A77-223356> (accessed 19th Feb 2013).
- PADMAN, R., MCCOLLEY, S. A., MILLER, D. P., KONSTAN, M. W., MORGAN, W. J., SCHECHTER, M. S., REN, C. L., WAGENER, J. S., INVESTIGATORS & COORDINATORS OF THE EPIDEMIOLOGIC STUDY OF CYSTIC, F. 2007. Infant care patterns at epidemiologic study of cystic fibrosis sites that achieve superior childhood lung function. *Pediatrics*, 119, e531-7.
- PAN, H. & COLE, T. 2002. User's guide to LmsGrowth. *London: Medical Research Council*, 2005.
- PANICO, L., BARTLEY, M., MARMOT, M., NAZROO, J. Y., SACKER, A. & KELLY, Y. J. 2007. Ethnic variation in childhood asthma and wheezing illnesses: findings from the Millennium Cohort Study. *Int J Epidemiol*, 36, 1093-102.
- PATTERSON, J. M., BUDD, J., GOETZ, D. & WARWICK, W. J. 1993. Family correlates of a 10-year pulmonary health trend in cystic fibrosis. *Pediatrics*, 91, 383-9.
- PAYNE, R. & ABEL, G. 2012. UK indices of multiple deprivation—a way to make comparisons across constituent countries easier. *Health statistics quarterly/Office for National Statistics*, 22.
- PENKETH, A. R., WISE, A., MEARNS, M. B., HODSON, M. E. & BATTEN, J. C. 1987. Cystic fibrosis in adolescents and adults. *Thorax*, 42, 526-32.
- PENNOCK, B. E., ROGERS, R. M. & MCCAFFREE, D. R. 1981. Changes in measured spirometric indices. What is significant? *Chest*, 80, 97-9.
- PETROU, S. & KUPEK, E. 2005. Socioeconomic differences in childhood hospital inpatient service utilisation and costs: prospective cohort study. *J Epidemiol Community Health*, 59, 591-7.
- PHAROAH, P. O., STEVENSON, C. J. & WEST, C. R. 2003. General Certificate of Secondary Education performance in very low birthweight infants. *Arch Dis Child*, 88, 295-8.
- PHELAN, P. & HEY, E. 1984. Cystic fibrosis mortality in England and Wales and in Victoria, Australia 1976-80. *Arch Dis Child*, 59, 71-3.
- PICKETT, K. E. & WILKINSON, R. G. 2007. Child wellbeing and income inequality in rich societies: ecological cross sectional study. *BMJ*, 335, 1080.
- POLLITT, R. J., DALTON, A., EVANS, S., HUGHES, H. N. & CURTIS, D. 1997. Neonatal screening for cystic fibrosis in the Trent region (UK): two-stage immunoreactive trypsin screening compared with a three-stage protocol with DNA analysis as an intermediate step. *J Med Screen*, 4, 23-8.
- POLLOCK, A. M., PRICE, D. & RODERICK, P. 2012a. Health and Social Care Bill 2011: a legal basis for charging and providing fewer health services to people in England. *BMJ*, 344, e1729.

- POLLOCK, A. M., PRICE, D., RODERICK, P., TREUHERZ, T., MCCOY, D., MCKEE, M. & REYNOLDS, L. 2012b. How the Health and Social Care Bill 2011 would end entitlement to comprehensive health care in England. *Lancet*, 379, 387-9.
- POST, P. N., WITTENBERG, J. & BURGERS, J. S. 2009. Do specialized centers and specialists produce better outcomes for patients with chronic diseases than primary care generalists? A systematic review. *Int J Qual Health Care*, 21, 387-96.
- POWER, C., MATTHEWS, S. & MANOR, O. 1996. Inequalities in self rated health in the 1958 birth cohort: lifetime social circumstances or social mobility? *BMJ*, 313, 449-53.
- PRESSLER, T. 2008. Review of recombinant human deoxyribonuclease (rhDNase) in the management of patients with cystic fibrosis. *Biologics*, 2, 611-7.
- QUINTON, H. B. & O'CONNOR, G. T. 2007. Current issues in quality improvement in cystic fibrosis. *Clin Chest Med*, 28, 459-72.
- QUITTNER, A. L., SCHECHTER, M. S., RASOULIYAN, L., HASELKORN, T., PASTA, D. J. & WAGENER, J. S. 2010. Impact of socioeconomic status, race, and ethnicity on quality of life in patients with cystic fibrosis in the United States. *Chest*, 137, 642-50.
- QUON, B. S. & GOSS, C. H. 2011. A story of success: continuous quality improvement in cystic fibrosis care in the USA. *Thorax*, 66, 1106-8.
- QUON, B. S., MAYER-HAMBLETT, N., AITKEN, M. L. & GOSS, C. H. 2012. Risk of post-lung transplant renal dysfunction in adults with cystic fibrosis. *Chest*, 142, 185-91.
- RABE, K. F., HURD, S., ANZUETO, A., BARNES, P. J., BUIST, S. A., CALVERLEY, P., FUKUCHI, Y., JENKINS, C., RODRIGUEZ-ROISIN, R., VAN WEEL, C., ZIELINSKI, J. & GLOBAL INITIATIVE FOR CHRONIC OBSTRUCTIVE LUNG, D. 2007. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med*, 176, 532-55.
- RABIN, H. R., BUTLER, S. M., WOHL, M. E., GELLER, D. E., COLIN, A. A., SCHIDLOW, D. V., JOHNSON, C. A., KONSTAN, M. W., REGELMANN, W. E. & EPIDEMIOLOGIC STUDY OF CYSTIC, F. 2004. Pulmonary exacerbations in cystic fibrosis. *Pediatr Pulmonol*, 37, 400-6.
- RAMSEY, B. W., BANKS-SCHLEGEL, S., ACCURSO, F. J., BOUCHER, R. C., CUTTING, G. R., ENGELHARDT, J. F., GUGGINO, W. B., KARP, C. L., KNOWLES, M. R., KOLLS, J. K., LIPUMA, J. J., LYNCH, S., MCCRAY, P. B., JR., RUBENSTEIN, R. C., SINGH, P. K., SORSCHER, E. & WELSH, M. 2012. Future directions in early cystic fibrosis lung disease research: an NHLBI workshop report. *Am J Respir Crit Care Med*, 185, 887-92.
- RAMSEY, B. W., DORKIN, H. L., EISENBERG, J. D., GIBSON, R. L., HARWOOD, I. R., KRAVITZ, R. M., SCHIDLOW, D. V., WILMOTT, R. W., ASTLEY, S. J., MCBURNIE, M. A. & ET AL. 1993. Efficacy of aerosolized tobramycin in patients with cystic fibrosis. *N Engl J Med*, 328, 1740-6.
- RANGANATHAN, S. C., PARSONS, F., GANGELL, C., BRENNAN, S., STICK, S. M., SLY, P. D. & AUSTRALIAN RESPIRATORY EARLY SURVEILLANCE TEAM FOR CYSTIC, F. 2011. Evolution of pulmonary inflammation and

- nutritional status in infants and young children with cystic fibrosis. *Thorax*, 66, 408-13.
- RASMUSSEN, J. N., GISLASON, G. H., RASMUSSEN, S., ABILDSTROM, S. Z., SCHRAMM, T. K., KOBER, L., DIDERICHSEN, F., OSLER, M., TORP-PEDERSEN, C. & MADSEN, M. 2007. Use of statins and beta-blockers after acute myocardial infarction according to income and education. *J Epidemiol Community Health*, 61, 1091-7.
- RASMUSSEN, J. N., RASMUSSEN, S., GISLASON, G. H., BUCH, P., ABILDSTROM, S. Z., KOBER, L., OSLER, M., DIDERICHSEN, F., TORP-PEDERSEN, C. & MADSEN, M. 2006. Mortality after acute myocardial infarction according to income and education. *J Epidemiol Community Health*, 60, 351-6.
- RAVALLION, M. 2001. Growth, inequality and poverty: looking beyond averages. *World development*, 29, 1803-1815.
- RICH-EDWARDS, J. W., KLEINMAN, K., MICHELS, K. B., STAMPFER, M. J., MANSON, J. E., REXRODE, K. M., HIBERT, E. N. & WILLETT, W. C. 2005. Longitudinal study of birth weight and adult body mass index in predicting risk of coronary heart disease and stroke in women. *BMJ*, 330, 1115.
- RIORDAN, J. R., ROMMENS, J. M., KEREM, B., ALON, N., ROZMAHEL, R., GRZELCZAK, Z., ZIELENSKI, J., LOK, S., PLAVSIC, N., CHOU, J. L. & ET AL. 1989. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science*, 245, 1066-73.
- ROBERTS, H. 2012. *What works in reducing inequalities in child health*, The Policy Press.
- ROBINSON, K. A., SALDANHA, I. J. & MCKOY, N. A. 2009. Management of infants with cystic fibrosis: a summary of the evidence for the cystic fibrosis foundation working group on care of infants with cystic fibrosis. *J Pediatr*, 155, S94-S105.
- ROMMENS, J. M., IANNUZZI, M. C., KEREM, B., DRUMM, M. L., MELMER, G., DEAN, M., ROZMAHEL, R., COLE, J. L., KENNEDY, D., HIDAKA, N. & ET AL. 1989. Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science*, 245, 1059-65.
- ROSENFELD, M., COATES, A., COREY, M., GOSS CH, HOWARD MB, MORGAN W, QUINTON H & M., S. 2005. Task Force to Evaluate Choice of Spirometric Reference Equations for the National Patient Registry: Summary and Recommendations. Cystic Fibrosis Foundation Registry Committee; Oct 1, 2005.
- ROSENFELD, M., DAVIS, R., FITZSIMMONS, S., PEPE, M. & RAMSEY, B. 1997. Gender gap in cystic fibrosis mortality. *Am J Epidemiol*, 145, 794-803.
- ROSENFELD, M., EMERSON, J., MCNAMARA, S., THOMPSON, V., RAMSEY, B. W., MORGAN, W. & GIBSON, R. L. 2012a. Risk factors for age at initial *Pseudomonas* acquisition in the cystic fibrosis epic observational cohort. *J Cyst Fibros*, 11, 446-53.
- ROSENFELD, M., RATJEN, F., BRUMBACK, L., DANIEL, S., ROWBOTHAM, R., MCNAMARA, S., JOHNSON, R., KRONMAL, R., DAVIS, S. D. & GROUP, I. S. 2012b. Inhaled hypertonic saline in infants and children younger than 6 years with cystic fibrosis: the ISIS randomized controlled trial. *JAMA*, 307, 2269-77.

- ROSENSTEIN, B. J. & CUTTING, G. R. 1998. The diagnosis of cystic fibrosis: a consensus statement. Cystic Fibrosis Foundation Consensus Panel. *J Pediatr*, 132, 589-95.
- ROSENTHAL, M. 2008. Annual assessment spirometry, plethysmography, and gas transfer in cystic fibrosis: do they predict death or transplantation. *Pediatr Pulmonol*, 43, 945-52.
- ROSENTHAL, M. 2009. How good are pulmonary function tests as an indicator of short and long term health status? . *Pediatr Pulmonol*, S32, 171-172.
- ROWE, S. M., MILLER, S. & SORSCHER, E. J. 2005. Cystic fibrosis. *N Engl J Med*, 352, 1992-2001.
- RUBIN, B. K. 1990. Exposure of children with cystic fibrosis to environmental tobacco smoke. *N Engl J Med*, 323, 782-8.
- RYAN, G., JAHNKE, N., REMMINGTON, T. & SMYTH, A. 2013. Inhaled antibiotics for pulmonary exacerbations in cystic fibrosis. *Paediatr Respir Rev*, 14, 27-8.
- RYAN, G., SINGH, M. & DWAN, K. 2011. Inhaled antibiotics for long-term therapy in cystic fibrosis. *Cochrane Database Syst Rev*, CD001021.
- SALDANA, P. S. & POMERANZ, J. L. 2012. Cystic fibrosis and the workplace: a review of the literature. *Work*, 42, 185-93.
- SALVATORE, D., BUZZETTI, R., BALDO, E., FORNERIS, M. P., LUCIDI, V., MANUNZA, D., MARINELLI, I., MESSORE, B., NERI, A. S., RAIA, V., FURNARI, M. L. & MASTELLA, G. 2010. An overview of international literature from cystic fibrosis registries 2. Neonatal screening and nutrition/growth. *J Cyst Fibros*, 9, 75-83.
- SALVATORE, D., BUZZETTI, R., BALDO, E., FORNERIS, M. P., LUCIDI, V., MANUNZA, D., MARINELLI, I., MESSORE, B., NERI, A. S., RAIA, V., FURNARI, M. L. & MASTELLA, G. 2011. An overview of international literature from cystic fibrosis registries. Part 3. Disease incidence, genotype/phenotype correlation, microbiology, pregnancy, clinical complications, lung transplantation, and miscellanea. *J Cyst Fibros*, 10, 71-85.
- SALVATORE, D., BUZZETTI, R., BALDO, E., FURNARI, M. L., LUCIDI, V., MANUNZA, D., MARINELLI, I., MESSORE, B., NERI, A. S., RAIA, V. & MASTELLA, G. 2012. An overview of international literature from cystic fibrosis registries. Part 4: update 2011. *J Cyst Fibros*, 11, 480-93.
- SANDERS, D. B., BITTNER, R. C., ROSENFELD, M., HOFFMAN, L. R., REDDING, G. J. & GOSS, C. H. 2010. Failure to recover to baseline pulmonary function after cystic fibrosis pulmonary exacerbation. *Am J Respir Crit Care Med*, 182, 627-32.
- SANDERS, D. B., BITTNER, R. C., ROSENFELD, M., REDDING, G. J. & GOSS, C. H. 2011. Pulmonary exacerbations are associated with subsequent FEV1 decline in both adults and children with cystic fibrosis. *Pediatr Pulmonol*, 46, 393-400.
- SAWICKI, G. S., SELLERS, D. E. & ROBINSON, W. M. 2009. High treatment burden in adults with cystic fibrosis: challenges to disease self-management. *J Cyst Fibros*, 8, 91-6.
- SAXENA, S., ELIAHOO, J. & MAJEED, A. 2002. Socioeconomic and ethnic group differences in self reported health status and use of health services by

- children and young people in England: cross sectional study. *BMJ*, 325, 520.
- SAXENA, S., MAJEED, A. & JONES, M. 1999. Socioeconomic differences in childhood consultation rates in general practice in England and Wales: prospective cohort study. *BMJ*, 318, 642-6.
- SCAMBLER, G. & HOPKINS, A. 1980. Social class, epileptic activity, and disadvantage at work. *J Epidemiol Community Health*, 34, 129-33.
- SCHECHTER, M. S. 2003. Non-genetic influences on cystic fibrosis lung disease: the role of sociodemographic characteristics, environmental exposures, and healthcare interventions. *Semin Respir Crit Care Med*, 24, 639-52.
- SCHECHTER, M. S. 2004. Non-genetic influences on CF lung disease: the role of sociodemographic characteristics, environmental exposures and healthcare interventions. *Pediatr Pulmonol Suppl*, 26, 82-5.
- SCHECHTER, M. S. 2011. Nongenetic influences on cystic fibrosis outcomes. *Curr Opin Pulm Med*, 17, 448-54.
- SCHECHTER, M. S. 2012. Benchmarking to improve the quality of cystic fibrosis care. *Curr Opin Pulm Med*, 18, 596-601.
- SCHECHTER, M. S. 2013. Wealth as a disease modifier in cystic fibrosis. *The Lancet Respiratory Medicine*, 1, 93-95.
- SCHECHTER, M. S. & GUTIERREZ, H. H. 2010. Improving the quality of care for patients with cystic fibrosis. *Curr Opin Pediatr*, 22, 296-301.
- SCHECHTER, M. S. & MARGOLIS, P. 2005. Improving subspecialty healthcare: lessons from cystic fibrosis. *J Pediatr*, 147, 295-301.
- SCHECHTER, M. S. & MARGOLIS, P. A. 1998. Relationship between socioeconomic status and disease severity in cystic fibrosis. *J Pediatr*, 132, 260-4.
- SCHECHTER, M. S., MCCOLLEY, S. A., REGELMANN, W., MILLAR, S. J., PASTA, D. J., WAGENER, J. S., KONSTAN, M. W. & MORGAN, W. J. 2011. Socioeconomic status and the likelihood of antibiotic treatment for signs and symptoms of pulmonary exacerbation in children with cystic fibrosis. *J Pediatr*, 159, 819-824 e1.
- SCHECHTER, M. S., MCCOLLEY, S. A., SILVA, S., HASELKORN, T., KONSTAN, M. W. & WAGENER, J. S. 2009. Association of socioeconomic status with the use of chronic therapies and healthcare utilization in children with cystic fibrosis. *J Pediatr*, 155, 634-9 e1-4.
- SCHECHTER, M. S., SHELTON, B. J., MARGOLIS, P. A. & FITZSIMMONS, S. C. 2001. The association of socioeconomic status with outcomes in cystic fibrosis patients in the United States. *Am J Respir Crit Care Med*, 163, 1331-7.
- SCHNEIDERMAN-WALKER, J., POLLOCK, S. L., COREY, M., WILKES, D. D., CANNY, G. J., PEDDER, L. & REISMAN, J. J. 2000. A randomized controlled trial of a 3-year home exercise program in cystic fibrosis. *J Pediatr*, 136, 304-10.
- SCHWARTZ, S. 1994. The fallacy of the ecological fallacy: the potential misuse of a concept and the consequences. *Am J Public Health*, 84, 819-24.
- SCOTET, V., DE BRAEKELEER, M., AUDREZET, M. P., QUERE, I., MERCIER, B., DUGUEPEROUX, I., ANDRIEUX, J., BLAYAU, M. & FEREC, C. 2002. Prenatal detection of cystic fibrosis by ultrasonography: a retrospective study of more than 346 000 pregnancies. *J Med Genet*, 39, 443-8.
- SELL, K., ZLOTNIK, S., NOONAN, K. & RUBIN, D. 2010. The Effect of Recession on Child Well-Being: A Synthesis of the Evidence by PolicyLab, The

Children's Hospital of Philadelphia  
[http://www.annarbor.com/Recession\\_ChildWellBeing\\_0-1.pdf](http://www.annarbor.com/Recession_ChildWellBeing_0-1.pdf) (accessed  
14th May 2013).

- SEMPLE, M. G., TAYLOR-ROBINSON, D. C., LANE, S. & SMYTH, R. L. 2011. Household tobacco smoke and admission weight predict severe bronchiolitis in infants independent of deprivation: prospective cohort study. *PLoS One*, 6, e22425.
- SEPHO 2012. Health Inequalities Gap Measurement Tool [http://www.sepho.nhs.uk/gap/gap\\_national.html](http://www.sepho.nhs.uk/gap/gap_national.html) (accessed 27th Feb 2013).
- SHOFF, S. M., AHN, H. Y., DAVIS, L., LAI, H. & WISCONSIN, C. F. N. S. G. 2006. Temporal associations among energy intake, plasma linoleic acid, and growth improvement in response to treatment initiation after diagnosis of cystic fibrosis. *Pediatrics*, 117, 391-400.
- SHUDY, M., DE ALMEIDA, M. L., LY, S., LANDON, C., GROFT, S., JENKINS, T. L. & NICHOLSON, C. E. 2006. Impact of pediatric critical illness and injury on families: a systematic literature review. *Pediatrics*, 118 Suppl 3, S203-18.
- SIMS, E. J., CLARK, A., MCCORMICK, J., MEHTA, G., CONNETT, G., MEHTA, A. & UNITED KINGDOM CYSTIC FIBROSIS DATABASE STEERING, C. 2007a. Cystic fibrosis diagnosed after 2 months of age leads to worse outcomes and requires more therapy. *Pediatrics*, 119, 19-28.
- SIMS, E. J., MCCORMICK, J., MEHTA, G., MEHTA, A. & COMMITTEE, U. C. D. S. 2005a. Newborn screening for cystic fibrosis is associated with reduced treatment intensity. *J Pediatr*, 147, 306-11.
- SIMS, E. J., MCCORMICK, J., MEHTA, G., MEHTA, A. & STEERING COMMITTEE OF THE, U. K. C. F. D. 2005b. Neonatal screening for cystic fibrosis is beneficial even in the context of modern treatment. *J Pediatr*, 147, S42-6.
- SIMS, E. J., MUGFORD, M., CLARK, A., AITKEN, D., MCCORMICK, J., MEHTA, G., MEHTA, A. & COMMITTEE, U. K. C. F. D. S. 2007b. Economic implications of newborn screening for cystic fibrosis: a cost of illness retrospective cohort study. *Lancet*, 369, 1187-95.
- SINGER, J. D. & WILLETT, J. B. 2003. *Applied longitudinal data analysis: modeling change and event occurrence: modeling change and event occurrence*, Oxford University Press, USA.
- SLY, P. D. & BRENNAN, S. 2004. Detecting early lung disease in cystic fibrosis: are current techniques sufficient? *Thorax*, 59, 1008-10.
- SLY, P. D., BRENNAN, S., GANGELL, C., DE KLERK, N., MURRAY, C., MOTT, L., STICK, S. M., ROBINSON, P. J., ROBERTSON, C. F., RANGANATHAN, S. C. & AUSTRALIAN RESPIRATORY EARLY SURVEILLANCE TEAM FOR CYSTIC, F. 2009. Lung disease at diagnosis in infants with cystic fibrosis detected by newborn screening. *Am J Respir Crit Care Med*, 180, 146-52.
- SMITH, G. D., HART, C., WATT, G., HOLE, D. & HAWTHORNE, V. 1998. Individual social class, area-based deprivation, cardiovascular disease risk factors, and mortality: the Renfrew and Paisley Study. *J Epidemiol Community Health*, 52, 399-405.
- SMITH, J. P. & GOLDMAN, D. 2010. Can patient self-management explain the health gradient? Goldman and Smith (2002) revisited: a response to Maitra. *Soc Sci Med*, 70, 813-5.

- SMYTH, A., O'HEA, U., FEYERABEND, C., LEWIS, S. & SMYTH, R. 2001. Trends in passive smoking in cystic fibrosis, 1993-1998. *Pediatr Pulmonol*, 31, 133-7.
- SMYTH, A., O'HEA, U., WILLIAMS, G., SMYTH, R. & HEAF, D. 1994. Passive smoking and impaired lung function in cystic fibrosis. *Arch Dis Child*, 71, 353-4.
- SMYTH, A. R. & WALTERS, S. 2012a. Prophylactic anti-staphylococcal antibiotics for cystic fibrosis. *Cochrane Database Syst Rev*, 12, CD001912.
- SMYTH, R. L. 2005. Diagnosis and management of cystic fibrosis. *Archives of disease in childhood-Education & practice edition*, 90, ep1-ep6.
- SMYTH, R. L. & WALTERS, S. 2012b. Oral calorie supplements for cystic fibrosis. *Cochrane Database Syst Rev*, 10, CD000406.
- SOUTHERN, K. W. 2004. Newborn screening for cystic fibrosis: the practical implications. *J R Soc Med*, 97 Suppl 44, 57-9.
- SOUTHERN, K. W., BARKER, P. M., SOLIS-MOYA, A. & PATEL, L. 2012. Macrolide antibiotics for cystic fibrosis. *Cochrane Database Syst Rev*, 11, CD002203.
- SOUTHERN, K. W., MERELLE, M. M., DANKERT-ROELSE, J. E. & NAGELKERKE, A. D. 2009. Newborn screening for cystic fibrosis. *Cochrane Database Syst Rev*, CD001402.
- SPENCER, N., BAMBANG, S., LOGAN, S. & GILL, L. 1999. Socioeconomic status and birth weight: comparison of an area-based measure with the Registrar General's social class. *J Epidemiol Community Health*, 53, 495-8.
- STANBROOK, M. B., COREY, M. & TULLIS, D. E. 2004. The repeatability of forced expiratory volume measurements in adults with cystic fibrosis. *Chest*, 125, 150-5.
- STANTON, M. C., TAYLOR-ROBINSON, D., HARRIS, D., PAIZE, F., MAKWANA, N., HACKETT, S. J., BAINES, P. B., RIORDAN, F. A., MARZOUK, O., THOMSON, A. P., DIGGLE, P. J., HART, C. A. & CARROL, E. D. 2011. Meningococcal disease in children in Merseyside, England: a 31 year descriptive study. *PLoS One*, 6, e25957.
- STEINKAMP, G. & WIEDEMANN, B. 2002. Relationship between nutritional status and lung function in cystic fibrosis: cross sectional and longitudinal analyses from the German CF quality assurance (CFQA) project. *Thorax*, 57, 596-601.
- STEPHENSON, A., HUX, J., TULLIS, E., AUSTIN, P. C., COREY, M. & RAY, J. 2010. Socioeconomic status and risk of hospitalization among individuals with cystic fibrosis in Ontario, Canada. *Pediatr Pulmonol*, 46, 376-384.
- STEPHENSON, A., HUX, J., TULLIS, E., AUSTIN, P. C., COREY, M. & RAY, J. 2011. Higher risk of hospitalization among females with cystic fibrosis. *J Cyst Fibros*, 10, 93-9.
- STERN, D. A., MORGAN, W. J., WRIGHT, A. L., GUERRA, S. & MARTINEZ, F. D. 2007. Poor airway function in early infancy and lung function by age 22 years: a non-selective longitudinal cohort study. *Lancet*, 370, 758-64.
- STERN, M., WIEDEMANN, B., WENZLAFF, P. & GERMAN CYSTIC FIBROSIS QUALITY ASSESSMENT, G. 2008. From registry to quality management: the German Cystic Fibrosis Quality Assessment project 1995 2006. *Eur Respir J*, 31, 29-35.
- STICK, S. M. 2009. The first 2 years of life: implications of recent findings. *Curr Opin Pulm Med*, 15, 615-20.

- STICK, S. M., BRENNAN, S., MURRAY, C., DOUGLAS, T., VON UNGERN-STERNBERG, B. S., GARRATT, L. W., GANGELL, C. L., DE KLERK, N., LINNANE, B., RANGANATHAN, S., ROBINSON, P., ROBERTSON, C., SLY, P. D. & AUSTRALIAN RESPIRATORY EARLY SURVEILLANCE TEAM FOR CYSTIC, F. 2009. Bronchiectasis in infants and preschool children diagnosed with cystic fibrosis after newborn screening. *J Pediatr*, 155, 623-8 e1.
- STIRBU, I., KUNST, A. E., MIELCK, A. & MACKENBACH, J. P. 2011. Inequalities in utilisation of general practitioner and specialist services in 9 European countries. *BMC Health Serv Res*, 11, 288.
- STRINGER, M. D. 2008. Biliary atresia: service delivery and outcomes. *Semin Pediatr Surg*, 17, 116-22.
- SUBRAMANIAN, S., CHEN, J., REHKOPF, D., WATERMAN, P. & KRIEGER, N. 2006. Subramanian et al. Respond to "Think Conceptually, Act Cautiously". *American Journal of Epidemiology*, 164, 841-844.
- SUFIAN, B. 2012. Access to government benefits or co-pay assistance increases access to coverage and care. *Pediatric Pulmonology*, Volume 47 S35
- TAYLOR-ROBINSON, D., AGARWAL, U., DIGGLE, P. J., PLATT, M. J., YOXALL, B. & ALFIREVIC, Z. 2011. Quantifying the impact of deprivation on preterm births: a retrospective cohort study. *PLoS One*, 6, e23163.
- TAYLOR-ROBINSON, D. & GOSLING, R. 2011. English north-south divide. Local authority budget cuts and health inequalities. *BMJ*, 342, d1487.
- TAYLOR-ROBINSON, D. & SCHECHTER, M. S. 2011. Health inequalities and cystic fibrosis. *BMJ*, 343, d4818.
- TAYLOR-ROBINSON, D., SMYTH, R. L., DIGGLE, P. & WHITEHEAD, M. 2013a. The effect of social deprivation on clinical outcomes and the use of treatments in the UK cystic fibrosis population: A longitudinal study *Lancet Respiratory Medicine*, 1, 121-128.
- TAYLOR-ROBINSON, D., WHITEHEAD, M., DIDERICHSEN, F., OLESEN, H. V., PRESSLER, T., SMYTH, R. L. & DIGGLE, P. 2012a. Understanding the natural progression in %FEV1 decline in patients with cystic fibrosis: a longitudinal study. *Thorax*, 67, 860-6.
- TAYLOR-ROBINSON, D., WHITEHEAD, M., DIDERICHSEN, F., VEBERT OLESEN, H., PRESSLER, T., SMYTH, R. & DIGGLE, P. 2013b. Author's response: understanding the natural progression in %FEV1 decline in patients with cystic fibrosis: a longitudinal study. *Thorax*, 68, 294-5.
- TAYLOR-ROBINSON, D., WHITEHEAD, M., SMYTH, R., DIGGLE, P., HENDERSON, R. & BARRETT, J. 2012b. WS3. 3 Longitudinal changes in lung function and risk of death in cystic fibrosis: developing a joint model for the UK population. *Journal of Cystic Fibrosis*, 11, S6.
- THE CYSTIC FIBROSIS GENOTYPE-PHENOTYPE CONSORTIUM 1993. Correlation between genotype and phenotype in patients with cystic fibrosis. The Cystic Fibrosis Genotype-Phenotype Consortium. *N Engl J Med*, 329, 1308-13.
- THEODORSON, T. 1995. Cardiovascular risk and risk reduction: a review of recent literature. *J Family Community Med*, 2, 19-26.
- THOMAS, C. L., O'ROURKE, P. K. & WAINWRIGHT, C. E. 2008. Clinical outcomes of Queensland children with cystic fibrosis: a comparison between tertiary centre and outreach services. *Med J Aust*, 188, 135-9.

- THORNTON, J., ELLIOTT, R. A., TULLY, M. P., DODD, M. & WEBB, A. K. 2005. Clinical and economic choices in the treatment of respiratory infections in cystic fibrosis: comparing hospital and home care. *J Cyst Fibros*, 4, 239-47.
- TWISK, J. W. 2003. *Applied longitudinal data analysis for epidemiology: a practical guide*, Cambridge University Press.
- TWISS, J., STEWART, A. W. & BYRNES, C. A. 2006. Longitudinal pulmonary function of childhood bronchiectasis and comparison with cystic fibrosis. *Thorax*, 61, 414-8.
- VAN AALDEREN, W. M., MANNES, G. P., BOSMA, E. S., ROORDA, R. J. & HEYMANS, H. S. 1995. Home care in cystic fibrosis patients. *Eur Respir J*, 8, 172-5.
- VAN DE MHEEN, H., STRONKS, K., LOOMAN, C. W. & MACKENBACH, J. P. 1998. Role of childhood health in the explanation of socioeconomic inequalities in early adult health. *J Epidemiol Community Health*, 52, 15-9.
- VAN DER DOEF, H. P., KOKKE, F. T., VAN DER ENT, C. K. & HOUWEN, R. H. 2011. Intestinal obstruction syndromes in cystic fibrosis: meconium ileus, distal intestinal obstruction syndrome, and constipation. *Curr Gastroenterol Rep*, 13, 265-70.
- VAN DER SCHANS, C., PRASAD, A. & MAIN, E. 2009. Chest physiotherapy compared to no chest physiotherapy for cystic fibrosis. *Cochrane Database of Systematic Reviews 2000, Issue 2. Art. No.: CD001401. DOI: 10.1002/14651858.CD001401.*
- VAN DIEMEN, C. C., POSTMA, D. S., SIEDLINSKI, M., BLOKSTRA, A., SMIT, H. A. & BOEZEN, H. M. 2011. Genetic variation in TIMP1 but not MMPs predict excess FEV1 decline in two general population-based cohorts. *Respir Res*, 12, 57.
- VAN DOORSLAER, E., MASSERIA, C. & KOOLMAN, X. 2006. Inequalities in access to medical care by income in developed countries. *CMAJ*, 174, 177-83.
- VAN KOOLWIJK, L. M., UITERWAAL, C. S., VAN DER LAAG, J., HOEKSTRA, J. H., GULMANS, V. A. & VAN DER ENT, C. K. 2002. Treatment of children with cystic fibrosis: central, local or both? *Acta Paediatr*, 91, 972-7; discussion 894-5.
- VANDENBRANDEN, S. L., MCMULLEN, A., SCHECHTER, M. S., PASTA, D. J., MICHAELIS, R. L., KONSTAN, M. W., WAGENER, J. S., MORGAN, W. J., MCCOLLEY, S. A., INVESTIGATORS & COORDINATORS OF THE EPIDEMIOLOGIC STUDY OF CYSTIC, F. 2012. Lung function decline from adolescence to young adulthood in cystic fibrosis. *Pediatr Pulmonol*, 47, 135-43.
- VANDEVANter, D., WAGENER, J., PASTA, D., ELKIN, E., JACOBS, J., MORGAN, W. & KONSTAN, M. 2010. Pulmonary outcome prediction (POP) tools for cystic fibrosis patients. *Pediatric pulmonology*, 45, 1156-116.
- VERMA, N., BUSH, A. & BUCHDAHL, R. 2005. Is there still a gender gap in cystic fibrosis? *Chest*, 128, 2824-34.
- VESTBO, J., SORENSEN, T., LANGE, P., BRIX, A., TORRE, P. & VISKUM, K. 1999. Long-term effect of inhaled budesonide in mild and moderate chronic obstructive pulmonary disease: a randomised controlled trial. *Lancet*, 353, 1819-23.
- VIVIANI, L., BOSSI, A., ASSAEL, B. M. & ITALIAN REGISTRY FOR CYSTIC FIBROSIS COLLABORATIVE, G. 2011. Absence of a gender gap in survival.

- An analysis of the Italian registry for cystic fibrosis in the paediatric age. *J Cyst Fibros*, 10, 313-7.
- WALD, N. J., MORRIS, J. K., RODECK, C. H., HADDOW, J. E. & PALOMAKI, G. E. 2003. Cystic fibrosis: selecting the prenatal screening strategy of choice. *Prenat Diagn*, 23, 474-83.
- WALDFOGEL, J. & WASHBROOK, E. 2010. Low income and early cognitive development in the UK <http://www.suttontrust.com/research/low-income-and-early-cognitive-development-in-the-uk/> (accessed 4th June 2013). *Sutton Trust*.
- WALKER, D. M., WEST, N. E., RAY, S. G. & BRITISH CARDIOVASCULAR SOCIETY WORKING GROUP ON ACUTE CARDIAC, C. 2012. From coronary care unit to acute cardiac care unit: the evolving role of specialist cardiac care. *Heart*, 98, 350-2.
- WALKER, S. P., WACHS, T. D., GRANTHAM-MCGREGOR, S., BLACK, M. M., NELSON, C. A., HUFFMAN, S. L., BAKER-HENNINGHAM, H., CHANG, S. M., HAMADANI, J. D., LOZOFF, B., GARDNER, J. M., POWELL, C. A., RAHMAN, A. & RICHTER, L. 2011. Inequality in early childhood: risk and protective factors for early child development. *Lancet*, 378, 1325-38.
- WALTERS, S., BRITTON, J. & HODSON, M. E. 1993. Demographic and social characteristics of adults with cystic fibrosis in the United Kingdom. *BMJ*, 306, 549-52.
- WALTERS, S., BRITTON, J. & HODSON, M. E. 1994. Hospital care for adults with cystic fibrosis: an overview and comparison between special cystic fibrosis clinics and general clinics using a patient questionnaire. *Thorax*, 49, 300-6.
- WANG, S. S., FITZSIMMONS, S. C., O'LEARY, L. A., ROCK, M. J., GWINN, M. L. & KHOURY, M. J. 2001. Early diagnosis of cystic fibrosis in the newborn period and risk of *Pseudomonas aeruginosa* acquisition in the first 10 years of life: A registry-based longitudinal study. *Pediatrics*, 107, 274-9.
- WANG, X., DOCKERY, D. W., WYPIJ, D., FAY, M. E. & FERRIS, B. G., JR. 1993. Pulmonary function between 6 and 18 years of age. *Pediatr Pulmonol*, 15, 75-88.
- WARK, P. & MCDONALD, V. M. 2009. Nebulised hypertonic saline for cystic fibrosis. *Cochrane Database Syst Rev*, CD001506.
- WAYLEN, A., STALLARD, N. & STEWART-BROWN, S. 2008. Parenting and health in mid-childhood: a longitudinal study. *Eur J Public Health*, 18, 300-5.
- WEINER, J. R., TOY, E. L., SACCO, P. & DUH, M. S. 2008. Costs, quality of life and treatment compliance associated with antibiotic therapies in patients with cystic fibrosis: a review of the literature. *Expert Opin Pharmacother*, 9, 751-66.
- WEISS, R. E. 2005. *Modeling longitudinal data*, Springer.
- WELKE, K. F., DIGGS, B. S., KARAMLOU, T. & UNGERLEIDER, R. M. 2008. The relationship between hospital surgical case volumes and mortality rates in pediatric cardiac surgery: a national sample, 1988-2005. *Ann Thorac Surg*, 86, 889-96; discussion 889-96.
- WEST, P. 1991. Rethinking the health selection explanation for health inequalities. *Soc Sci Med*, 32, 373-84.
- WHEELER, B. W. & BEN-SHLOMO, Y. 2005. Environmental equity, air quality, socioeconomic status, and respiratory health: a linkage analysis of

- routine data from the Health Survey for England. *J Epidemiol Community Health*, 59, 948-54.
- WHITEHEAD, M. 1990. *The concepts and principles of equity and health*. WHO 1990, Regional Office for Europe: Copenhagen, WHO.
- WHITEHEAD, M. 1992. The concepts and principles of equity and health. *International journal of health services*, 22, 429-445.
- WHITEHEAD, M. & DAHLGREN, G. 2007. Concepts and principles for tackling social inequities in health: Levelling up Part 1. *Copenhagen: WHO*.
- WHITEHEAD, M., DAHLGREN, G. & EVANS, T. 2001. Equity and health sector reforms: can low-income countries escape the medical poverty trap? *Lancet*, 358, 833-6.
- WHITEHEAD, M., DIDERICHSEN, F. & BURSTROM, B. 2000. Researching the impact of public policy on inequalities in health. *Understanding Health Inequalities*. Open University Press, Buckingham.
- WHITEHEAD, M. & POPAY, J. 2010. Swimming upstream? Taking action on the social determinants of health inequalities. *Soc Sci Med*, 71, 1234-6; discussion 1254-8.
- WHO 1977. Manual of mortality analysis. Geneva, World Health Organization (Division of Health Statistics) [http://whqlibdoc.who.int/publications/17009\\_\(chI-chVI\).pdf](http://whqlibdoc.who.int/publications/17009_(chI-chVI).pdf) (accessed 4th June 2013).
- WHO 1981. *Global strategy for health for all by the year 2000* <http://whqlibdoc.who.int/publications/9241800038.pdf> (accessed 24th May 2013), World Health Organization.
- WHO 2007. A conceptual framework for action on the social determinants of health. Discussion paper for the Commission on Social Determinants of Health [http://www.who.int/sdhconference/resources/ConceptualframeworkforactiononSDH\\_eng.pdf](http://www.who.int/sdhconference/resources/ConceptualframeworkforactiononSDH_eng.pdf) (accessed 28th Feb 2013).
- WIJLAARS, L. P., JOHNSON, L., VAN JAARSVELD, C. H. & WARDLE, J. 2011. Socioeconomic status and weight gain in early infancy. *Int J Obes (Lond)*, 35, 963-70.
- WILKINSON, R. G. 1992. Income distribution and life expectancy. *BMJ*, 304, 165-8.
- WILKINSON, R. G. 2005. *The impact of inequality: how to make sick societies healthier*, London, Routledge.
- WILKINSON, R. G. & PICKETT, K. 2010. *The spirit level: why more equal societies almost always do better*, Allen Lane.
- WINTHROP, A. L., BRASEL, K. J., STAHOVIC, L., PAULSON, J., SCHNEEBERGER, B. & KUHN, E. M. 2005. Quality of life and functional outcome after pediatric trauma. *J Trauma*, 58, 468-73; discussion 473-4.
- WOHL, A. S. 1983. *Endangered lives: public health in Victorian Britain*, JM Dent and Sons Ltd.
- WOLFENDEN, L. L. & SCHECHTER, M. S. 2009. Genetic and non-genetic determinants of outcomes in cystic fibrosis. *Paediatr Respir Rev*, 10, 32-6.
- YOHANNES, A. M., WILLGOSS, T. G., FATOYE, F. A., DIP, M. D. & WEBB, K. 2012. Relationship between anxiety, depression, and quality of life in adult patients with cystic fibrosis. *Respir Care*, 57, 550-6.

- ZEITLIN, P. L. 2008. Cystic fibrosis and estrogens: a perfect storm. *J Clin Invest*, 118, 3841-4.
- ZEMEL, B. S., JAWAD, A. F., FITZSIMMONS, S. & STALLINGS, V. A. 2000. Longitudinal relationship among growth, nutritional status, and pulmonary function in children with cystic fibrosis: analysis of the Cystic Fibrosis Foundation National CF Patient Registry. *J Pediatr*, 137, 374-80.
- ZIAIAN, T., SAWYER, M. G., REYNOLDS, K. E., CARBONE, J. A., CLARK, J. J., BAGHURST, P. A., COUPER, J. J., KENNEDY, D., MARTIN, A. J., STAUGAS, R. E. & FRENCH, D. J. 2006. Treatment burden and health-related quality of life of children with diabetes, cystic fibrosis and asthma. *J Paediatr Child Health*, 42, 596-600.

## Appendix 3 (Pertaining to Chapter 3, Methods)



24 September 2008

Medical Research Council  
20 Park Crescent  
London  
W1B 1AL

Dear Sir / Madam

**Re: Dr David Taylor-Robinson's MRC Population Health Scientist Fellowship application: The effect of socioeconomic status on outcomes for people with cystic fibrosis: A longitudinal study**

I confirm that the Cystic Fibrosis Trust fully supports David's MRC fellowship application and we look forward to continued collaboration with him on this project.

David's project is important and of great relevance to the work of the CF Trust. One of the Trust's objectives is to ensure appropriate clinical services for all patients with cystic fibrosis. The proposed project will highlight and attempt to explain any differences in clinical outcomes and healthcare access by socioeconomic status. Such work will inform policy, and assist the Trust as it advocates for a universal standard of healthcare for all patients with cystic fibrosis.

The Cystic Fibrosis Trust maintains and manages the Cystic Fibrosis Trust Patient Registry, which is the main data source for the proposed fellowship. The CF Trust Database committee has considered a draft of this proposal, and fully supports the use of the registry data for this project. Professor Rosalind Smyth is one of David's supervisors. Ros is well known to the Trust, and has an international reputation for her research on Cystic Fibrosis. David has already undertaken preliminary analyses using a subset of the data, with the support of the CF Registry manager.

As the UK's only national charity dedicated to all aspects of Cystic Fibrosis, and a major funder of CF research, the Trust is well placed to provide support and advice on this project.

Yours sincerely

A handwritten signature in black ink that reads 'Alan Larsen'.

Alan Larsen  
Director of Research & Finance

see off **cf**

11 London Road, Bromley, Kent BR1 1BY  
Tel: 020 8464 7211 • Fax: 020 8313 0472 • [www.cftrust.org.uk](http://www.cftrust.org.uk)

Patron: HRH Princess Alexandra, the Hon. Lady Ogilvy, KG, GCVO President: Duncan Black CBE  
Cystic Fibrosis Trust registered as a charity number 1079049  
A company limited by guarantee registered in England and Wales number 3880213  
Registered office: 11 London Road, Bromley, Kent BR1 1BY



Københavns Universitet  
Institut for Folkesundhedsvidenskab  
Øster Farimagsgade 5  
1014 København K

Att. Finn Diderichsen  
Sendt til: fidi@sund.ku.dk

**30. august 2010**

### Udtalelse i forbindelse med anmeldelse af "Cystisk Fibrose Kohorte"

Datatilsynet  
Borgergade 28, 5.  
1300 København K

CVR-nr. 11-88-37-  
29

Telefon 3319 3200  
Fax 3319 3218

E-post  
dt@datatilsynet.  
dk  
www.datatilsynet.  
dk

J.nr. 2010-54-  
0989

Sagsbehandler  
Camilla Daasnes  
Direkte 3319 3225

Ifølge persondatalovens<sup>1</sup> § 45, stk. 1, nr. 3, skal der forinden iværksættelse af behandling, som udelukkende finder sted i videnskabeligt eller statistisk øjemed, og som er omfattet af anmeldelsespligten i § 43, indhentes en udtalelse fra Datatilsynet.

Datatilsynet har den 2. august 2010 modtaget ovennævnte anmeldelse, hvoraf fremgår, at Københavns Universitet er dataansvarlig myndighed. Det fremgår endvidere, at behandlingen udelukkende skal finde sted i videnskabeligt eller statistisk øjemed. Den 4. august 2010 har Datatilsynet modtaget diverse ændringer til anmeldelsen.

Anmeldelsen giver ikke tilsynet anledning til bemærkninger.

Datatilsynet skal dog henlede opmærksomheden på følgende: Efter persondataloven § 10, stk. 1, kan der behandles oplysninger som nævnt i lovens § 7, stk. 1, eller § 8, hvis dette alene sker med henblik på at udføre statistiske eller videnskabelige undersøgelser af væsentlig samfundsmæssig betydning, og hvis behandlingen er nødvendig for udførelsen af undersøgelsen.

Af § 10, stk. 2, fremgår endvidere, at oplysninger omfattet af § 10, stk. 1, ikke senere må behandles i andet end statistisk eller videnskabeligt øjemed. Det samme gælder de ikke-følsomme oplysninger, som indgår i behandlingen, jf. § 10, stk. 2, 2. pkt. Oplysninger om en person må således ikke efterfølgende anvendes til at træffe afgørelser eller foranstaltninger vedrørende den pågældende. Oplysningerne må endvidere kun videregives til tredjemand efter forudgående tilladelse fra Datatilsynet, og i givet fald kun med henblik på udførelse af undersøgelser i statistisk eller videnskabeligt øjemed, jf. § 10, stk. 3.

Endelig skal Datatilsynet særligt gøre opmærksom på, at den dataansvarlige til enhver tid skal sikre sig, at dokumenter og andre præsentationer, som publiceres eller på anden måde gøres tilgængelig for andre på internettet, usb-nøgle eller på andet elektronisk medie, ikke indeholder personoplysninger.

<sup>1</sup> Lov nr. 429 af 31. maj 2000 om behandling af personoplysninger med senere ændringer.

Der skal vises særlig agtpågivenhed i forbindelse med brug af grafiske præsentationer i Excel og PowerPoint, da de uforvarende kan indeholde indlejrede persondata i form af regneark, tabeller mv. Præsentationer, der gøres tilgængelig på internettet, bør derfor omformateres til Portable Digital Format (PDF), da dette fjerner eventuelle indlejrede Excel-tabeller.

Eventuelle ændringer af de forhold, som er omtalt i anmeldelsesblanketten, skal anmeldes til Datatilsynet efter reglen i persondatalovens § 46.

Anmeldelsen offentliggøres i fortegnelsen på Datatilsynets hjemmeside.

Med venlig hilsen

Camilla Daasnes

## Port CF screens: Demographic data

  <b>FULL POSTCODE</b>
<input type="text"/> <input type="radio"/> foreign <input type="radio"/> unknown

 <b>GP POSTCODE</b>
<input type="text"/>

  <b>BIOLOGICAL PARENTAL HEIGHT</b>
Mother's height <input type="text"/> <input type="radio"/> cm <input type="radio"/> in <input type="checkbox"/> Unknown
Father's height <input type="text"/> <input type="radio"/> cm <input type="radio"/> in <input type="checkbox"/> Unknown
<i>Information not required for patients 16 and above.</i>

  <b>GENOTYPE</b>
<input type="checkbox"/> Genotype information confirmed
Has this patient been genotyped? <input type="radio"/> Yes <input type="radio"/> No
Date <input type="text"/> <input type="checkbox"/> Date Unknown
Mutation 1 <input type="text" value="Select"/> <input type="button" value="v"/>
if other, specify <input type="text"/> re-enter <input type="text"/>
Mutation 2 <input type="text" value="Select"/> <input type="button" value="v"/>
if other, specify <input type="text"/> re-enter <input type="text"/>
for data entry - the caret (^) is to be used in place of a delta ( $\Delta$ ) or the word "delta"

  <b>COMPLICATIONS AT BIRTH</b>
<input type="radio"/> none
<input type="radio"/> meconium ileus/other intestinal obstruction managed medically
<input type="radio"/> meconium ileus/other intestinal obstruction managed surgically
<input type="radio"/> unknown

**IDENTIFYING PATIENT DEMOGRAPHIC DATA**

CF Patient Number : B166217  
 Care Centre: Test Centre - Matt  
 Centre Patient ID : B79496

Last Name

Last Name at Birth (if different)

First Name

County Of Birth

Date Of Birth (dd/mm/yyyy)

**GENDER**

Male  Female

**RACE (Check all that apply)**

Caucasian  
 Black African  
 Black Caribbean  
 Black Other  
 Indian  
 Pakistani  
 Bangladeshi  
 Chinese  
 Asian Other  
 Mixed Race  
 Individual preferred not to answer  
 Clinician preferred not to ask question  
 Other (Please Specify)

**NATIONAL HEALTH SERVICE NUMBER / CHI NUMBER**

NHS Number   refused  unassigned/foreign  unknown

CHI Number   refused  unassigned/foreign  unknown

**DEATH**

Date of Death   Date is an Estimate

Primary Cause of Death

if other, specify:

**CF DIAGNOSIS**

Date of diagnosis (may be earlier than first sweat test.)

**Diagnosis suggested by/symptoms at first CF work-up (select all that apply):**

Acute or persistent respiratory symptoms

Oedema

Electrolyte imbalance

Failure to thrive/malnutrition

Family history

Genotype

Liver problems

Meconium ileus/other intestinal obstruction

Nasal polyps/sinus disease

Prenatal screening (CVS, amnio)

Neonatal screening

Rectal prolapse

Steatorrhea/abnormal stools/malabsorption

Unknown

Other (specify)

Chloride Sweat Test Value   Unknown

CF diagnosis reversed during year

If reversed, select reason diagnosis was reversed

DNA Analysis

Transepithelial Potential Differences

Repeat Normal Sweat Testing

Other

**Encounter, General:**

<b>Information for Patient ID: B166217</b>
<b>Patient Name:</b> M1 BAKER
<b>Birth date:</b> 01/01/2000
<b>Centre Patient ID:</b> B79496

This encounter is part of  

**Functions performed at this encounter**

- |   |  |
|---|--|
| <input checked="" type="checkbox"/> Nutrition     | <input checked="" type="checkbox"/> Respiratory Microbiology |
| <input checked="" type="checkbox"/> Laboratory    | <input checked="" type="checkbox"/> Pulmonary Assessment     |
| <input checked="" type="checkbox"/> Complications | <input checked="" type="checkbox"/> Pulmonary Therapies      |

<b>Encounter Date:</b> <input type="text" value="01/01/2008"/>
--

  <b>PATIENT STATUS</b>
At the time of clinical visit, the patient was: <input type="radio"/> well <input type="radio"/> unwell <input type="radio"/> unknown

  <b>HEIGHT</b>
<input type="text"/> <input type="checkbox"/> not measured
Measure <input type="text"/> <input type="radio"/> cm. <input type="radio"/> in. percentile:

  <b>WEIGHT</b>
<input type="text"/> <input type="checkbox"/> not measured
Measure <input type="text"/> <input type="radio"/> kg. <input type="radio"/> lb. percentile:

 <b>SOCIAL WORKER CONSULTATION</b>
<input type="checkbox"/> Patient Consulted with a Social Worker at this visit

**Encounter, nutrition:**

  <b>NUTRITIONAL ASSESSMENT</b>	
<input type="checkbox"/>	Patient was seen by a Dietitian
Assessment of oral intake	
<input type="radio"/>	Done
<input type="radio"/>	Not Done
Is patient currently receiving supplemental feeding? <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Unknown	
If Yes, indicate Feeding Route:	
<input type="checkbox"/>	oral supplementation
<input type="checkbox"/>	nasogastric tube
<input type="checkbox"/>	gastrostomy tube/button
<input type="checkbox"/>	jejunal tube
<input type="checkbox"/>	total parenteral nutrition
Is patient currently receiving pancreatic enzyme supplements?	
<input type="radio"/>	Yes
<input type="radio"/>	No
<input type="radio"/>	Unknown
<input type="checkbox"/>	Creon 5000 scoop
<input type="checkbox"/>	Creon 8000
<input type="checkbox"/>	Creon 10000
<input type="checkbox"/>	Creon 25000
<input type="checkbox"/>	Creon 40000
<input type="checkbox"/>	Pancrease 5000
<input type="checkbox"/>	Nutrizyme
<input type="checkbox"/>	Other, Please specify <input type="text"/>
Number per day <input type="text"/>	
Acid Blocker (Daily use. Check all that apply since last visit)	
<input type="checkbox"/>	Antacids
<input type="checkbox"/>	H2 Blocker
<input type="checkbox"/>	Proton Pump Inhibitor
<input type="checkbox"/>	None
<input type="checkbox"/>	Unknown

### Encounter, pulmonary assessments:

 <b>PULMONARY FUNCTION TESTS (PFTs)</b>		
If no values given, select reason		
<input type="radio"/> Not Done		
<input type="radio"/> Unable to perform reliable test		
FVC	measure <input type="text"/> (L)	% predicted:
FEV1	measure <input type="text"/> (L)	% predicted:
FEF25-75	measure <input type="text"/> (L)	% predicted:

 <b>PULMONARY ASSESSMENT</b>
Was the patient seen by a doctor at this visit?
<input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Unknown
Based on the assessment of the care team at this visit, was the patient experiencing an increase in respiratory symptoms or a pulmonary exacerbation?
<input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Unknown
If the patient was experiencing an increase in respiratory symptoms or a pulmonary exacerbation at this visit (i.e., the question above was answered yes), then select the treatment option below that best matches the prescribed treatment plan.
Select <input type="text"/> 
(Specify) <input type="text"/>

**Encounter, pulmonary therapies:**

  <b>PULMONARY THERAPIES</b>		
<input type="checkbox"/> Pulmonary Therapies confirmed		
<b>Airway Clearance Techniques</b> (check only one primary means of airway clearance and all secondary forms of airway clearance that apply)		
<u>Technique</u>	Primary (select one)	Secondary (select all that apply)
Positive Expiratory Pressure (PEP)	<input type="radio"/>	<input type="checkbox"/>
Postural drainage with clapping	<input type="radio"/>	<input type="checkbox"/>
Forced expiration techniques (eg, autogenic drainage, huff cough, active cycle breathing)	<input type="radio"/>	<input type="checkbox"/>
Oscillating PEP (eg, Flutter, acapella, IPV)	<input type="radio"/>	<input type="checkbox"/>
High frequency chest wall compression (eg, Vest)	<input type="radio"/>	<input type="checkbox"/>
Exercise	<input type="radio"/>	<input type="checkbox"/>
Other (specify) <input type="text"/>	<input type="radio"/>	<input type="checkbox"/>
None of the above	<input type="radio"/>	
<b>Pulmonary Medications used since last encounter</b> (check all that applied as of last clinical encounter)		
<b>Mucolytics</b>		
<input type="checkbox"/> DNase Frequency: <input type="text" value="Select"/>		
<input type="checkbox"/> Acetylcysteine		
<b>Inhaled antibiotics</b>		
<input type="checkbox"/> Tobramycin solution for inhalation Frequency: <input type="text" value="Select"/>		
<input type="checkbox"/> Other aminoglycoside Frequency: <input type="text" value="Select"/>		
<input type="checkbox"/> Colistin Frequency: <input type="text" value="Select"/>		
<input type="checkbox"/> Promixin Frequency: <input type="text" value="Select"/>		
<input type="checkbox"/> Chronic macrolide antibiotic <input type="text" value="Select"/>		

Chronic oral antibiotic (i.e. not prescribed to treat an exacerbation)

Check all that apply

- Quinolone
- Cephalosporin
- Amoxicillin
- Tetracycline
- Flucloxacillin
- Cotrimoxazole
- Other

High-dose ibuprofen (e.g. 25-30 mg/kg)

(Total mg/dose)  mg

Hypertonic saline

Concentration %  Frequency

*Bronchodilators (oral)*

- Beta agonist
- Theophylline product

*Bronchodilators (inhaled)*

- Short acting beta agonist
- Long acting beta agonist
- Short acting anticholinergic
- Long acting anticholinergic
- Combination beta agonist and anticholinergic

*Corticosteroids*

- Oral
- Inhaled
- Inhaled in combination with a bronchodilator

Leukotriene modifiers

Mast cell stabilizers

Antifungals

*Drug Intolerance*

- DNase
- Tobramycin solution for inhalation
- Colistin
- Macrolide antibiotics
- High-dose ibuprofen
- Hypertonic saline
- IV antibiotics, Please Specify :

This patient is not on any of the above pulmonary medications

## Encounter, respiratory microbiology:

 <b>RESPIRATORY MICROBIOLOGY</b>	
If culture was performed, where was it performed?	
<input type="radio"/> At CF Centre/clinic	<input type="radio"/> Elsewhere
Date of culture	
<input type="text"/>	
<u>Type of Culture</u>	
<input type="checkbox"/> sputum	
<input type="checkbox"/> throat/nasal	
<input type="checkbox"/> bronchoscopy	
<u>Culture Result</u>	
<input type="checkbox"/> no growth/sterile culture	
<input type="checkbox"/> normal flora	
<input type="checkbox"/> <i>Pseudomonas aeruginosa</i>	
<input type="checkbox"/> mucoid	
<input type="checkbox"/> non-mucoid	
<input type="checkbox"/> unknown	
<input type="checkbox"/> <i>Burkholderia cepacia</i> complex	
<input type="checkbox"/> <i>B. Cenocepacia</i>	
<input type="checkbox"/> <i>B. Multivorans</i>	
<input type="checkbox"/> Other <i>Burkholderia Cepacia</i>	
<input type="checkbox"/> <i>Stenotrophomonas (Xanthomonas) maltophilia</i>	
<input type="checkbox"/> Other <i>pseudomonas</i> species	
<input type="checkbox"/> <i>Staphylococcus aureus</i>	
<input type="checkbox"/> MRSA (methicillin resistant <i>Staph aureus</i> )	
<input type="checkbox"/> <i>Haemophilus influenzae</i> (any species)	
<input type="checkbox"/> <i>Aspergillus</i> (any species)	
<input type="checkbox"/> non-tuberculous mycobacterium	
<input type="checkbox"/> <i>Escherichia coli</i> ( <i>E coli</i> )	
<input type="checkbox"/> <i>Klebsiella</i> (any species)	
<input type="checkbox"/> Other gram negative (e.g. <i>Burkholderia gladioli</i> is NOT included in <i>B. cepacia</i> complex)	
<input type="checkbox"/> <i>Alcaligenes (Achromobacter) xylosoxidans</i>	
<input type="checkbox"/> <i>Pandoria</i>	
<input type="checkbox"/> Other, Specify <input type="text"/>	

**Encounter, laboratory:**

  <b>SERUM CREATININE</b>	
<input type="checkbox"/> not done	
Level	<input type="text"/> mmol/dl

  <b>LIVER ENZYMES DRAWN</b>	
<input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Unknown	

  <b>CFRD SCREENING</b>			
Was this patient screened for CFRD?			
<input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> No-patient has known CFRD <input type="radio"/> Unknown			
		<b>if OGTT performed</b>	
random blood glucose	<input type="text"/> mmol/litre	fasting blood glucose	<input type="text"/> mmol/litre
fasting blood glucose	<input type="text"/> mmol/litre	2 hour	<input type="text"/> mmol/litre

**Encounter, complications:**

  <b>COMPLICATIONS (select all that apply)</b>	
<input type="checkbox"/> complications confirmed	
<input type="checkbox"/> Allergic Bronchial Pulmonary Aspergillosis (ABPA)	<input type="checkbox"/> Absence of Vas Deferens
<input type="checkbox"/> Arthritis	<input type="checkbox"/> Arthropathy
<input type="checkbox"/> Asthma	<input type="checkbox"/> Atyp. Mycobact. Inf. (req. Rx)
<input type="checkbox"/> Bone fracture	<input type="checkbox"/> Cancer confirmed by histology
<input type="checkbox"/> CFRD	<input type="checkbox"/> Chronic Pseudomonas aeruginosa
<input type="checkbox"/> Chronic Staph Aureus	<input type="checkbox"/> Cirrhosis with portal hypertension
<input type="checkbox"/> Cirrhosis with no portal hypertension	<input type="checkbox"/> Depression
<input type="checkbox"/> Dist Int Obst Synd (DIOS)	<input type="checkbox"/> Fib colonopathy/colonic strict.(report incidence only)
<input type="checkbox"/> Gallbladder Dis req surgery	<input type="checkbox"/> GERD (Gastro-Esoph. Ref. Dis.)
<input type="checkbox"/> GI Bleed req hosp non variceal	<input type="checkbox"/> GI Bleed req hosp variceal
<input type="checkbox"/> Hearing loss	<input type="checkbox"/> Hemoptysis, massive
<input type="checkbox"/> Hypertension	<input type="checkbox"/> Kidney Stones
<input type="checkbox"/> Liver disease, non-cirrhosis	<input type="checkbox"/> Liver enzymes elevated
<input type="checkbox"/> Nasal polyps req surgery	<input type="checkbox"/> NTM
<input type="checkbox"/> Osteopenia	<input type="checkbox"/> Osteoporosis
<input type="checkbox"/> Pancreatitis	<input type="checkbox"/> Peptic ulcer disease
<input type="checkbox"/> Pneumothorax req chest tube	<input type="checkbox"/> Port inserted or replaced
<input type="checkbox"/> Rectal prolapse	<input type="checkbox"/> Renal failure req dialysis
<input type="checkbox"/> Sinus Disease (symptomatic)	
Other: <input type="text"/>	

## Annual Survey Data

### Information for Patient ID: B166217

Patient Name: M1 BAKER

Birth date: 01/01/2000

Center Patient ID: B79496

### PATIENT STATUS

Since the last annual review, this patient was:

- Seen at this Centre/Clinic
- Transferred to another Centre/Clinic (seen or not seen)

Patient's date of death was

(Note: If you enter a date of death, please go to the Demographics section to update cause of death.)

Patient's Full Postcode at year end was   foreign  unknown

confirm full postcode

 **PULMONARY**

Since the last annual review, has the patient had any of the following:

Oxygen therapy?

Select 

Used non-invasive ventilation?

- Yes  No  Unknown

Had a Chest X Ray performed?  Yes  No  Unknown

Had a Dexa scan performed?  Normal  Abnormal  Not Done

Had Liver USS performed?  Normal  Abnormal  Not Done

Does the patient grow Pseudomonas?  Yes  No  Unknown

If yes, was this pseudomonas?

- Intermittent  Chronic

Does the patient grow Staph.aureus?  Yes  No  Unknown

If yes, was this Staph.aureus?

- Intermittent  Chronic

Did the patient receive an influenza vaccination?

- Yes  No  Unknown

Did the patient receive a pneumovax vaccination?  Yes  No  Unknown

Did the patient smoke cigarettes? Select 

If female then 'Oestrogen' will replace 'Testosterone' in the growth and nutrition section.

**GROWTH AND NUTRITION**

Does this patient take pancreatic enzyme supplements?  Yes  No  Unknown

Fat soluble vitamin levels measured?  
 Yes  No  Unknown

Has this Patient been on Testosterone?  
 Yes  No  Unknown

Was the patient prescribed daily Ursodeoxycholic acid (e.g. 20 mg/kg)?  
 Yes  No  Unknown

**UPDATE ON CFRD STATUS**

Not Done  
 Normal Glucose Metabolism (includes normal, random, fasting, or OGTT)  
 Impaired Glucose Tolerance (FBG 6.1-6.9, 2-h PG 7.8-11.0 mmol/L)

CF-related diabetes  
 CF-related diabetes without fasting hyperglycemia (FBG <6.0, 2-h PG >= 11.1 mmol/L)  
 CF-related diabetes with fasting hyperglycemia (FBG >=6.0 mmol/L)

Hgb A1C value   Not Done  Unknown

Retinopathy  Yes  No  Unknown

Microalbuminuria  Yes  No  Unknown

Was the patient prescribed treatment for CFRD?  Yes  No

If yes, select all that apply

Dietary change  
 Oral hypoglycemic agents  
 Intermittent Insulin (with illness, steroids, etc.)  
 Chronic requiring Insulin

 **TRANSPLANTATION**

Since the last annual review, has the patient been evaluated for a transplant

- Yes    No    Unknown

If Yes, were they

- accepted  
 declined  
 deferred

Did the patient receive a transplant?  Yes    No    Unknown

If yes, how many transplants did the patient receive?

- 1    2    3 or more

Which type(s) of transplant(s) did the patient receive? (select all that apply)

- Lung: Bilateral  
 Lung: Heart/Lung  
 Lung: Lobar/cadaveric  
 Lung: Lobar/living donor  
 Liver  
 Other

*specify type of transplant*

If lung transplant received, indicate month and day of first lung transplant.

Were there post transplant complications?  Yes    No

If yes, select those that apply below:

- Bronchiolitis obliterans syndrome  
 Lympho-proliferative disorder  
 Atypical infection  
 Renal failure  
 Other (specify below)

**CLINICAL TRIALS**

Since the last annual review, has this patient participated in any ethically approved protocols?

Yes    No    Unknown

If Yes, select those that apply below:

CF Trust Funded Studies

In house studies

Commercial Studies

Other (specify)

**IV DAYS**

Since the last annual review:

Hospital IV days

Home IV days

**SOCIO-ECONOMIC STATUS**

*Educational Levels*

Patient:

Mother of Patient:

Father of Patient:

If the patient is 16 or older the following will appear as socio-economic status:

? SOCIO-ECONOMIC STATUS	
<i>Educational Levels</i>	
Patient:	Unknown/Not Applicable ▼
Mother of Patient:	Unknown/Not Applicable ▼
Father of Patient:	University ▼
Spouse of Patient/Partner:	Unknown/Not Applicable ▼

? AGE 16 AND OLDER	
Marital Status	Divorced ▼
Employment (Select all that apply)	
<input checked="" type="checkbox"/>	Full time
<input type="checkbox"/>	Part Time
<input type="checkbox"/>	Full Time Homemaker
<input type="checkbox"/>	Student
<input type="checkbox"/>	Unemployed
<input type="checkbox"/>	Disabled
<input type="checkbox"/>	Retired
<input type="checkbox"/>	Unknown

If patient is female and older than xxxx the following will appear in the annual survey:

? PREGNANCY	
Since the last annual review, has this patient been pregnant? <input type="radio"/> Yes <input type="radio"/> No <input type="radio"/> Unknown	
If Yes, indicate outcome:	Select ▼

## Appendix 4 (Pertaining to Chapter 4)

The presentation of the model tables for clinical outcomes follows the same basic format. The baseline model is presented first, adjusted for baseline covariates, then deprivation is added (final model – the main results are from these models), and then other covariates of potential interest are added (final\_plus). See Figure 24, Chapter 4 for further explanation.

### Weight

**Table 20: Final regression models for weight SD score in <18 age group**

	<i>baseline</i>	<i>final</i>	<i>final_plus</i>
Constant	-0.97609*** (0.08326)	-0.76679*** (0.09124)	-0.99648*** (0.10440)
age	0.20961*** (0.01460)	0.17231*** (0.01769)	0.25400*** (0.02594)
age2 (split line effect)	-0.22797*** (0.01516)	-0.18983*** (0.01838)	-0.26711*** (0.02732)
Number of F508 alleles: 0/2	0.06410 (0.10624)	0.06228 (0.10612)	0.12097 (0.10659)
Number of F508 alleles: 1/2	0.02148 (0.06189)	0.01447 (0.06184)	0.05747 (0.06245)
Male	0.02713 (0.05728)	0.03527 (0.05724)	0.03353 (0.05718)
nonwhite	0.25721 (0.13949)	0.33507* (0.14012)	0.32258* (0.14001)
Screened	0.27766*** (0.06434)	0.28633*** (0.06427)	0.31266*** (0.06441)
Age x Number of F508 alleles: 0/2	-0.04467 (0.03089)	-0.04279 (0.03088)	-0.06150* (0.03102)
Age x Number of F508 alleles: 1/2	-0.02603 (0.01686)	-0.02331 (0.01687)	-0.03841* (0.01716)
Age x male	0.06098*** (0.01566)	0.05831*** (0.01568)	0.06010*** (0.01565)
age x nonwhite	-0.15834*** (0.04014)	-0.17607*** (0.04040)	-0.17066*** (0.04036)
Age x screened	-0.06142*** (0.01666)	-0.06239*** (0.01666)	-0.07167*** (0.01675)
age2 x Number of F508 alleles: 0/2	0.04677 (0.03200)	0.04497 (0.03199)	0.06298* (0.03213)
age2 x Number of F508 alleles: 1/2	0.03860* (0.01754)	0.03611* (0.01755)	0.05105** (0.01784)
age2 x male	-0.07615*** (0.01634)	-0.07335*** (0.01636)	-0.07568*** (0.01634)
age2 x nonwhite	0.14914*** (0.04227)	0.16692*** (0.04254)	0.16188*** (0.04249)
age2 x screened	0.05920*** (0.01769)	0.06005*** (0.01769)	0.06914*** (0.01779)
Deprivation score		-0.00932*** (0.00169)	-0.00949*** (0.00169)
Age x Deprivation score		0.00168*** (0.00045)	0.00174*** (0.00045)
Age2 x Deprivation score		-0.00173*** (0.00047)	-0.00179*** (0.00047)
<i>Pseudomonas</i> colonisation			0.42669*** (0.10238)
CFRD			-1.99523***

Pancreatic insufficiency			(0.56705) 0.21250***
Age x <i>Pseudomonas</i> colonisation			(0.05401) -0.12952***
Age x CFRD			(0.03593) 0.80867***
Age x Pancreatic insufficiency			(0.19694) -0.08365***
age2 x <i>Pseudomonas</i> colonisation			(0.02015) 0.12339***
age2 x CFRD			(0.03666) -0.84934***
age2 x Pancreatic insufficiency			(0.19948) 0.08199***
Log-likelihood	-	-	(0.02145)
Deviance	25550.15779	25528.81159	25475.60363
AIC	51100.31558	51057.62317	50951.20726
BIC	51204.31558	51167.62317	51079.20726
N	51635.03792	51623.19488	51609.32706
Groups	29235	29235	29235
	5775	5775	5775

\*P < 0.05, \*\* P < 0.01, \*\*\* P < 0.001

Standard errors in parentheses, birthyear coefficients not shown

The deprivation effect is multiplied by 58 to generate the contrast between the mid point of the least and most deprived quintile  
age2 is the coefficient for the split line at age three in the weight analysis

**Table 21: Final regression models for weight SD score in >18 age group**

	<i>baseline</i>	<i>final</i>	<i>final_plus</i>
Constant	-0.92805*** (0.09886)	-0.81353*** (0.10244)	-0.72702*** (0.10417)
Age-18	0.01834*** (0.00390)	0.01839*** (0.00391)	0.01688*** (0.00445)
Number of F508 alleles: 0/2	0.02131 (0.06528)	0.02906 (0.06518)	0.01587 (0.06498)
Number of F508 alleles: 1/2	0.22333*** (0.04983)	0.22908*** (0.04973)	0.21717*** (0.04953)
Male	-0.33035*** (0.04727)	-0.33225*** (0.04714)	-0.33590*** (0.04684)
Non-white	-0.59068*** (0.14073)	-0.54262*** (0.14092)	-0.54116*** (0.14011)
Screened	0.06358 (0.07901)	0.06145 (0.07886)	0.05919 (0.07848)
Age-18 x Number of F508 alleles: 0/2	0.01222 (0.00633)	0.01236 (0.00633)	0.01280* (0.00633)
Age-18 x Number of F508 alleles: 1/2	0.00123 (0.00446)	0.00123 (0.00446)	0.00131 (0.00446)
Age-18 x male	0.01059* (0.00425)	0.01053* (0.00425)	0.01072* (0.00423)
Age-18 x nonwhite	0.01448 (0.01430)	0.01385 (0.01431)	0.01372 (0.01426)
Age-18 x screened	-0.01288 (0.00897)	-0.01328 (0.00897)	-0.01223 (0.00895)
Deprivation score		-0.00537*** (0.00130)	-0.00538*** (0.00129)
<i>Pseudomonas</i> colonisation			-0.06166*** (0.01563)
CFRD			-0.07284** (0.02509)
Pancreatic insufficiency			-0.04669 (0.02421)
Age-18 x <i>Pseudomonas</i> colonisation			0.00292 (0.00158)
Age-18 x CFRD			0.00439 (0.00229)
Age-18 x Pancreatic insufficiency			-0.00042 (0.00233)
Log-likelihood	-18655.24983	-18646.75337	-18619.45472
Deviance	37310.49966	37293.50675	37238.90944
AIC	37410.49966	37395.50675	37352.90944
BIC	37806.46524	37799.39164	37804.31020
N	20319	20319	20319
Groups	4041	4041	4041

\*P < 0.05, \*\* P < 0.01, \*\*\* P < 0.001

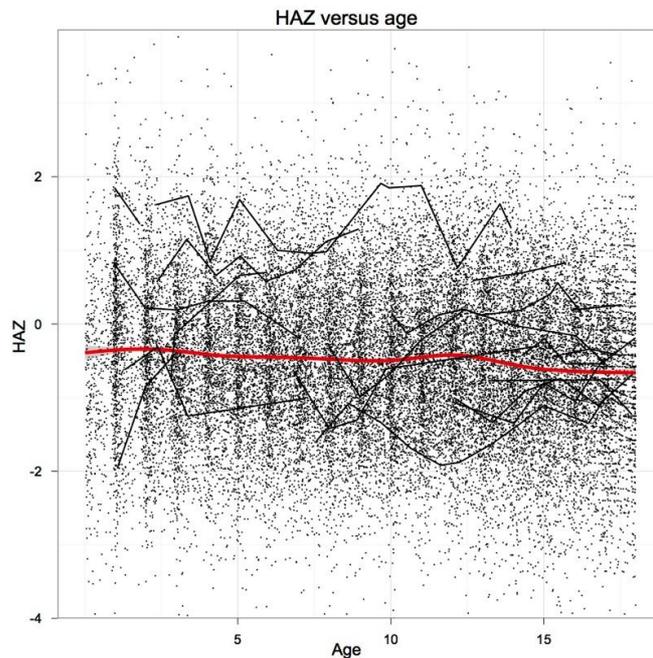
Standard errors in parentheses, birth year coefficients not shown

The deprivation effect is multiplied by 58 to generate the contrast between the mid point of the least and most deprived quintile

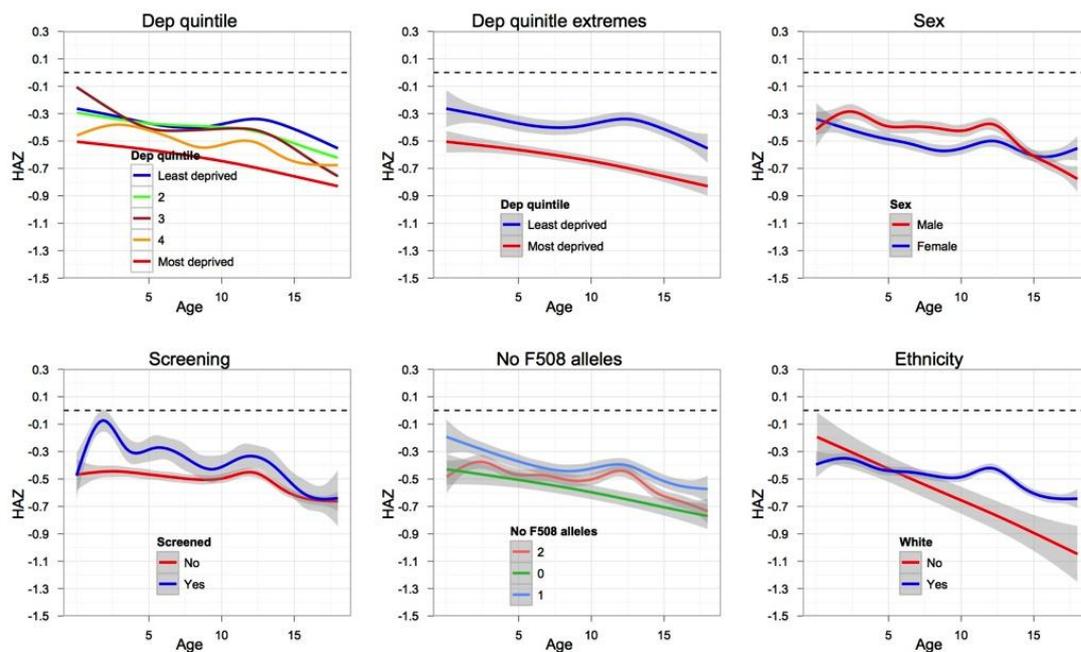
## Height

**Figure 83: Spaghetti plot of height for age SD score versus age in paediatric age group.**

*Mean smoother in red.*

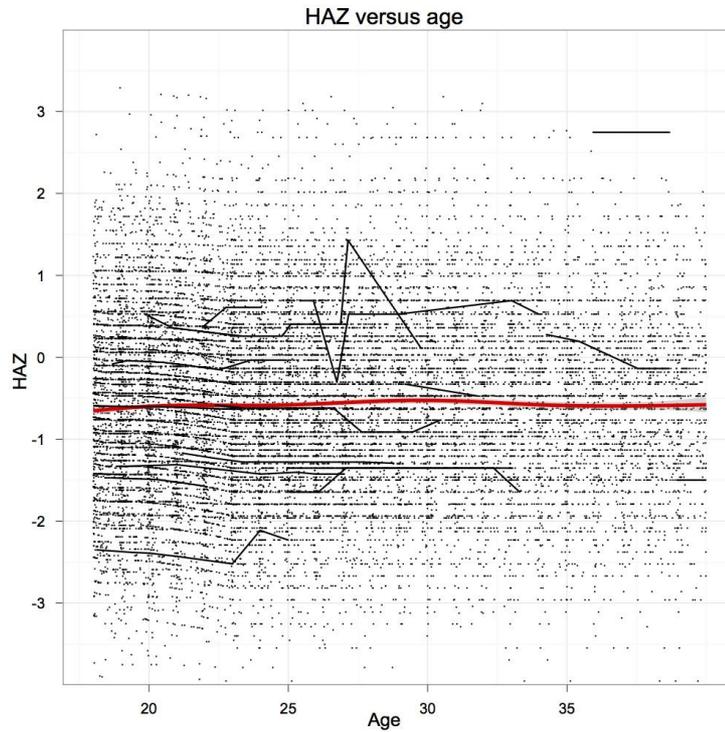


**Figure 84: Exploratory analysis showing smoothed means of height for age Z score versus age, stratified by covariates**

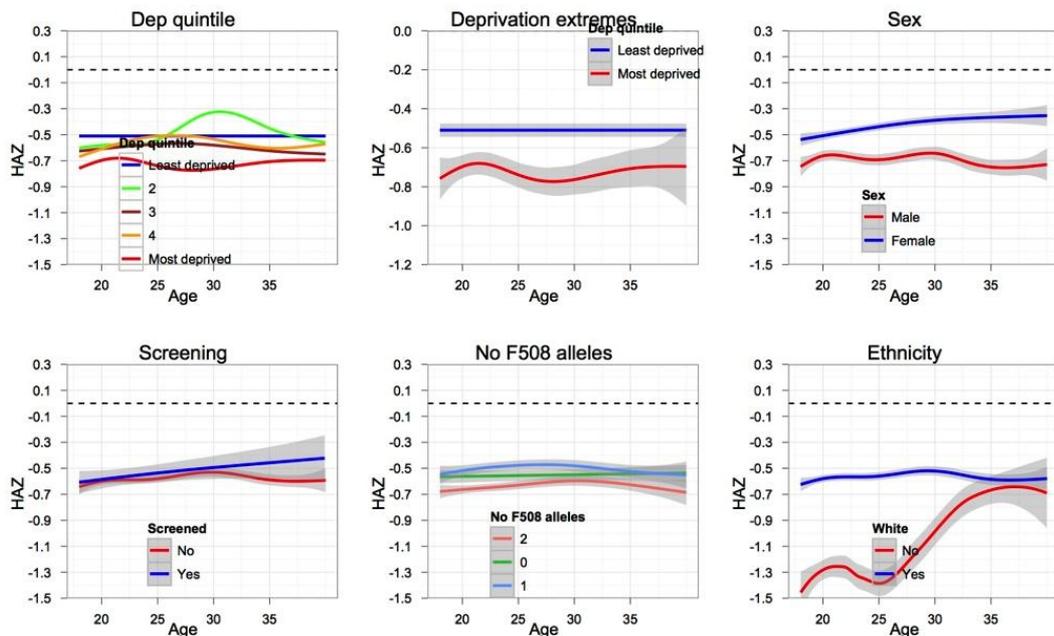


**Figure 85: Spaghetti plot of height for age Z score versus age in adult age group.**

*Mean smoother in red.*



**Figure 86: Exploratory analysis showing smoothed means of height for age Z score versus age, in adult age group, stratified by covariates**



**Table 22: Final regression models for height SD score in <18 age group**

	<i>baseline</i>	<i>final</i>	<i>final_plus</i>
Constant	-0.81565*** (0.07383)	-0.69634*** (0.07615)	-0.75521*** (0.08023)
Age	0.01171*** (0.00296)	0.01154*** (0.00296)	0.01899*** (0.00368)
Number of F508 alleles: 0/2	-0.06081 (0.07076)	-0.05825 (0.07068)	-0.04948 (0.07080)
Number of F508 alleles: 1/2	-0.00061 (0.04436)	-0.00004 (0.04431)	0.00507 (0.04436)
Male	0.17876*** (0.04099)	0.17969*** (0.04094)	0.18414*** (0.04089)
Non-white	0.20212* (0.10059)	0.23514* (0.10062)	0.24153* (0.10047)
Screened	0.27200*** (0.05060)	0.27879*** (0.05056)	0.28415*** (0.05050)
Age x Number of F508 alleles: 0/2	0.00172 (0.00529)	0.00184 (0.00529)	0.00113 (0.00529)
Age x Number of F508 alleles: 1/2	0.00562 (0.00348)	0.00582 (0.00347)	0.00538 (0.00347)
Age x male	-0.01225*** (0.00332)	-0.01218*** (0.00331)	-0.01287*** (0.00329)
Age x non-white	-0.02714*** (0.00805)	-0.02756*** (0.00804)	-0.02747*** (0.00802)
Age x screened	-0.01691*** (0.00407)	-0.01710*** (0.00407)	-0.01787*** (0.00405)
Deprivation score		-0.00526*** (0.00086)	-0.00527*** (0.00086)
<i>Pseudomonas</i> colonisation			0.06316** (0.02094)
CFRD			0.18810* (0.09121)
Pancreatic insufficiency			0.02859 (0.02637)
Age x <i>Pseudomonas</i> colonisation			-0.00578*** (0.00175)
Age x CFRD			-0.02129*** (0.00620)
Age x Pancreatic insufficiency			-0.00450 (0.00235)
Log-likelihood	-22676.57348	-22657.81459	-22622.62490
Deviance	45353.14697	45315.62917	45245.24980
AIC	45445.14697	45409.62917	45351.24980
BIC	45825.77234	45798.52902	45789.79644
N	28983	28983	28983
Groups	5750	5750	5750

\*P < 0.05, \*\* P < 0.01, \*\*\* P < 0.001

Standard errors in parentheses, birth year coefficients not shown

The deprivation effect is multiplied by 58 to generate the contrast between the mid point of the least and most deprived quintile

**Table 23: Final regression models for height SD score in >18 age group**

	<i>baseline</i>	<i>final</i>	<i>final_plus</i>
Constant	-0.34842*** (0.08095)	-0.22411** (0.08435)	-0.23732** (0.08488)
Age-18	-0.00509*** (0.00140)	-0.00508*** (0.00140)	-0.00376* (0.00167)
Number of F508 alleles: 0/2	0.11255** (0.03632)	0.11796** (0.03626)	0.11780** (0.03627)
Number of F508 alleles: 1/2	0.09951** (0.03236)	0.10350** (0.03227)	0.10411** (0.03228)
Male	-0.23973*** (0.03487)	-0.24207*** (0.03474)	-0.24325*** (0.03473)
Non-white	-0.40991*** (0.08556)	-0.38022*** (0.08558)	-0.38249*** (0.08553)
Screened	-0.07661 (0.04345)	-0.07743 (0.04337)	-0.07910 (0.04336)
Age-18 x Number of F508 alleles: 0/2	-0.00197 (0.00238)	-0.00191 (0.00238)	-0.00214 (0.00239)
Age-18 x Number of F508 alleles: 1/2	0.00006 (0.00165)	0.00006 (0.00165)	-0.00012 (0.00165)
Age-18 x male	0.00019 (0.00155)	0.00018 (0.00155)	0.00031 (0.00155)
Age-18 x non-white	-0.00109 (0.00528)	-0.00119 (0.00528)	-0.00110 (0.00527)
Age-18 x screened	-0.00333 (0.00336)	-0.00347 (0.00336)	-0.00327 (0.00336)
Deprivation score		-0.00530*** (0.00105)	-0.00529*** (0.00105)
<i>Pseudomonas</i> colonisation			0.01101 (0.00674)
CFRD			-0.02267* (0.01079)
Pancreatic insufficiency			0.01229 (0.01046)
Age-18 x <i>Pseudomonas</i> colonisation			-0.00035 (0.00068)
Age-18 x CFRD			0.00180 (0.00096)
Age-18 x Pancreatic insufficiency			-0.00158 (0.00099)
Log-likelihood	-3722.83080	-3710.19476	-3704.59670
Deviance	7445.66160	7420.38951	7409.19340
AIC	7545.66160	7522.38951	7523.19340
BIC	7941.48919	7926.13366	7974.43686
N	20263	20263	20263
Groups	4046	4046	4046

\*P < 0.05, \*\* P < 0.01, \*\*\* P < 0.001

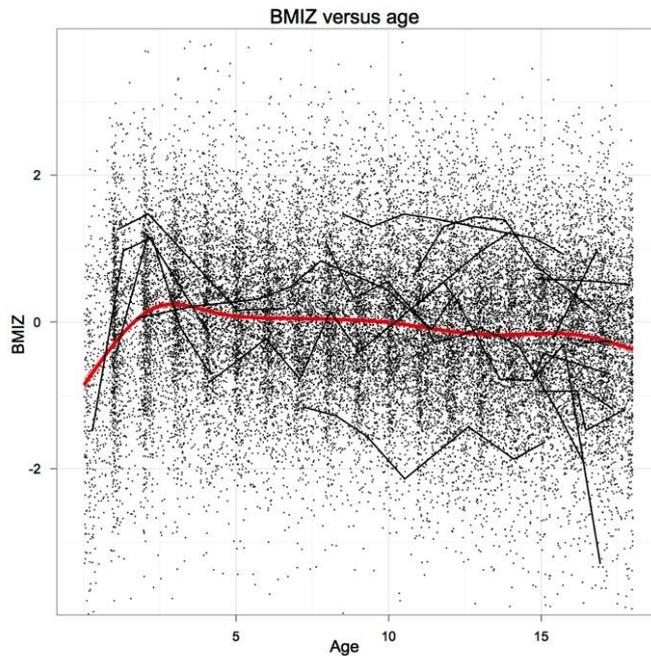
Standard errors in parentheses, birth year coefficients not shown

The deprivation effect is multiplied by 58 to generate the contrast between the mid point of the least and most deprived quintile

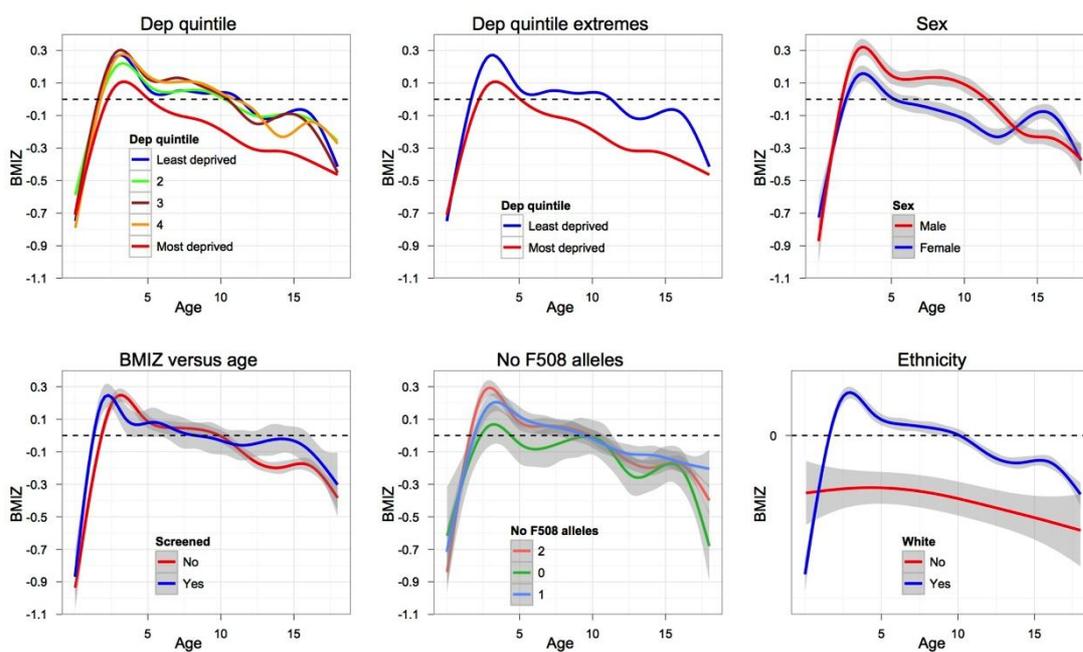
## BMI

**Figure 87: Spaghetti plot of BMI for age Z score versus age in paediatric age group.**

*Mean smoother in red.*

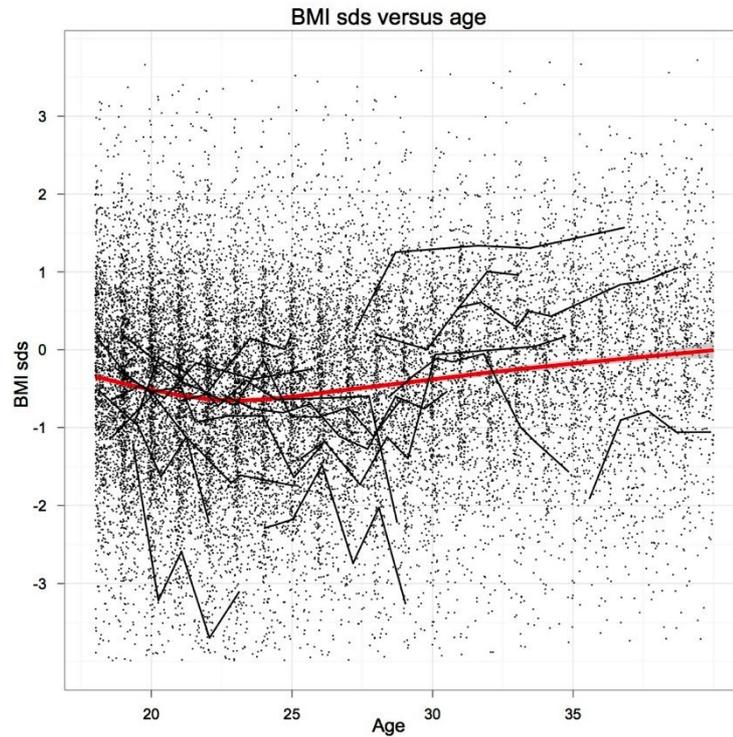


**Figure 88: Exploratory analysis showing smoothed means of BMI for age Z score versus age, stratified by covariates**

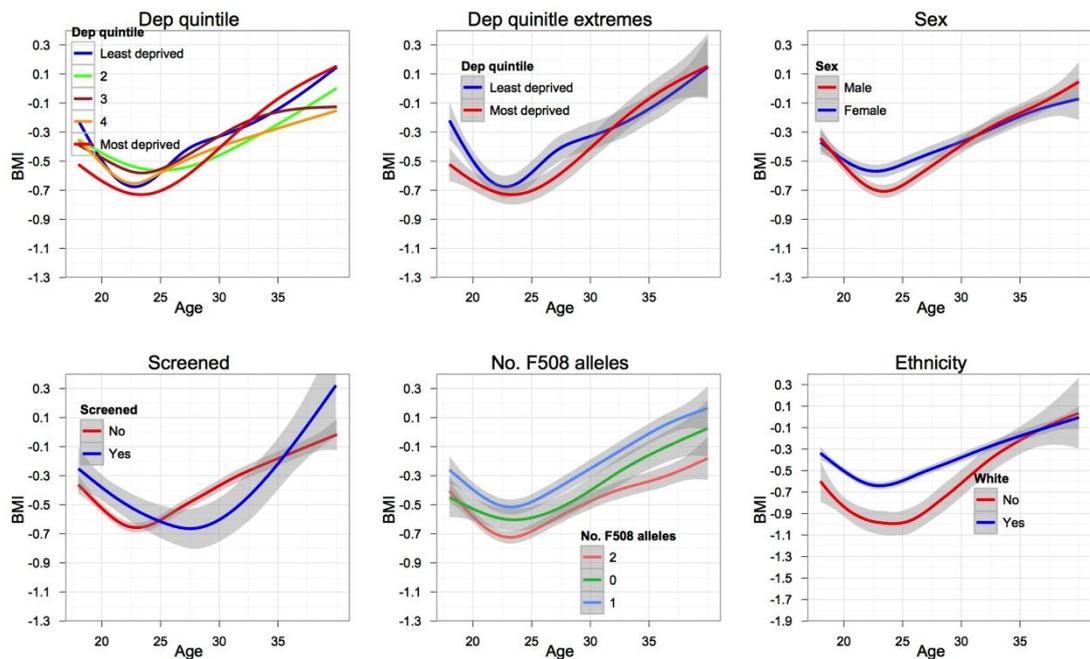


**Figure 89: Spaghetti plot of BMI for age Z score versus age in adult age group.**

*Mean smoother in red.*



**Figure 90: Exploratory analysis showing smoothed means of BMI for age Z score versus age, in adult age group, stratified by covariates**



**Table 24: Final regression models for BMI SD score in <18 age group.**

	<i>baseline</i>	<i>final</i>	<i>final_plus</i>
Constant	-0.53677*** (0.09942)	-0.48286*** (0.10110)	-0.72782*** (0.12972)
age	0.15859*** (0.02777)	0.15798*** (0.02777)	0.26685*** (0.04093)
age2	-0.17618*** (0.02879)	-0.17570*** (0.02880)	-0.28421*** (0.04268)
Number of F508 alleles: 0/2	-0.03001 (0.15024)	-0.03069 (0.15020)	0.04315 (0.15239)
Number of F508 alleles: 1/2	-0.21116* (0.08859)	-0.21288* (0.08857)	-0.16805 (0.08997)
Male	-0.09253 (0.08161)	-0.09166 (0.08159)	-0.09316 (0.08170)
Non-white	-0.04359 (0.18742)	-0.02379 (0.18750)	-0.03729 (0.18782)
Screened	0.09886 (0.08736)	0.10080 (0.08734)	0.12893 (0.08783)
Age x Number of F508 alleles: 0/2	0.00079 (0.05455)	0.00141 (0.05455)	-0.02586 (0.05519)
Age x Number of F508 alleles: 1/2	0.05595 (0.03192)	0.05660 (0.03192)	0.03889 (0.03236)
Age x male	0.11392*** (0.02947)	0.11384*** (0.02947)	0.11436*** (0.02947)
Age x non-white	-0.11929 (0.07002)	-0.12099 (0.07002)	-0.11714 (0.07004)
Age x screened	-0.03121 (0.03215)	-0.03067 (0.03215)	-0.04176 (0.03230)
age2 x Number of F508 alleles: 0/2	0.00089 (0.05688)	0.00036 (0.05688)	0.02685 (0.05750)
age2 x Number of F508 alleles: 1/2	-0.04799 (0.03332)	-0.04850 (0.03332)	-0.03104 (0.03376)
age2 x male	-0.13391*** (0.03079)	-0.13381*** (0.03079)	-0.13414*** (0.03079)
age2 x non-white	0.13173 (0.07400)	0.13337 (0.07400)	0.13029 (0.07400)
age2 x screened	0.03530 (0.03406)	0.03464 (0.03407)	0.04640 (0.03420)
Deprivation score		-0.00229** (0.00079)	-0.00229** (0.00079)
<i>Pseudomonas</i> colonisation			0.29458* (0.14939)
CFRD			-0.73942 (0.89849)
Pancreatic insufficiency			0.25291** (0.08457)
Age x <i>Pseudomonas</i> colonisation			-0.10559* (0.05176)
Age x CFRD			0.31656 (0.30612)
Age x Pancreatic insufficiency			-0.10866*** (0.03139)
age2 x <i>Pseudomonas</i> colonisation			0.10529* (0.05251)
age2 x CFRD			-0.32927 (0.30825)
age2 x Pancreatic insufficiency			0.10736** (0.03291)
Log-likelihood	-28162.34270	-28158.12899	-28131.32222
Deviance	56324.68539	56316.25798	56262.64444
AIC	56434.68539	56428.25798	56392.64444
BIC	56889.77526	56891.62221	56930.47792
N	28980	28980	28980
Groups	5745	5745	5745

\*P < 0.05, \*\* P < 0.01, \*\*\* P < 0.001

Standard errors in parentheses, birth year coefficients not shown

The deprivation effect is multiplied by 58 to generate the contrast between the mid point of the least and most deprived quintile.

Age2 is the coefficient for the split line at age three

**Table 25: Final regression models for BMI SD score in >18 age group.**

	<i>baseline</i>	<i>final</i>	<i>final_plus</i>
Constant	-0.21805* (0.09144)	-0.17173 (0.09519)	-0.04328 (0.10561)
Age-18	-0.07687*** (0.01117)	-0.07677*** (0.01117)	-0.08689*** (0.01768)
age2	0.10160*** (0.01333)	0.10152*** (0.01333)	0.11180*** (0.02018)
Number of F508 alleles: 0/2	-0.06500 (0.07590)	-0.06111 (0.07592)	-0.08529 (0.07637)
Number of F508 alleles: 1/2	0.17795*** (0.05348)	0.18013*** (0.05347)	0.16327** (0.05346)
Male	0.00526 (0.04935)	0.00446 (0.04934)	-0.00124 (0.04908)
Non-white	-0.29930* (0.15130)	-0.27837 (0.15174)	-0.26552 (0.15084)
Screened	0.04868 (0.08262)	0.04777 (0.08260)	0.04648 (0.08210)
Age-18 x Number of F508 alleles: 0/2	0.02560 (0.02020)	0.02537 (0.02019)	0.02874 (0.02046)
Age-18 x Number of F508 alleles: 1/2	-0.00979 (0.01363)	-0.00980 (0.01363)	-0.00837 (0.01374)
Age-18 x male	-0.03936** (0.01252)	-0.03934** (0.01252)	-0.03862** (0.01254)
Age-18 x non-white	0.03474 (0.03951)	0.03412 (0.03951)	0.02954 (0.03952)
Age-18 x screened	0.02545 (0.02234)	0.02532 (0.02234)	0.02815 (0.02236)
age2 x Number of F508 alleles: 0/2	-0.01458 (0.02384)	-0.01423 (0.02384)	-0.01810 (0.02412)
age2 x Number of F508 alleles: 1/2	0.01498 (0.01618)	0.01497 (0.01618)	0.01297 (0.01633)
age2 x male	0.06502*** (0.01495)	0.06495*** (0.01495)	0.06441*** (0.01497)
age2 x non-white	-0.02736 (0.04853)	-0.02708 (0.04852)	-0.02163 (0.04855)
age2 x screened	-0.05167 (0.02884)	-0.05189 (0.02884)	-0.05556 (0.02887)
Deprivation score		-0.00203 (0.00117)	-0.00202 (0.00116)
<i>Pseudomonas</i> colonisation			-0.07043* (0.03202)
CFRD			-0.16322** (0.05294)
Pancreatic insufficiency			-0.07341 (0.05079)
Age-18 x <i>Pseudomonas</i> colonisation			0.00154 (0.00984)
Age-18 x CFRD			0.04219** (0.01520)
Age-18 x Pancreatic insufficiency			0.00480 (0.01517)
age2 x <i>Pseudomonas</i> colonisation			0.00147 (0.01105)
age2 x CFRD			-0.04591** (0.01662)
age2 x Pancreatic insufficiency			-0.00577 (0.01681)
Log-likelihood	-19862.44785	-19860.93596	-19827.07408
Deviance	39724.89570	39721.87192	39654.14815
AIC	39842.89570	39841.87192	39792.14815
BIC	40309.06561	40315.94301	40337.32991
N	19954	19954	19954
Groups	4029	4029	4029

\* $P < 0.05$ , \*\*  $P < 0.01$ , \*\*\*  $P < 0.001$

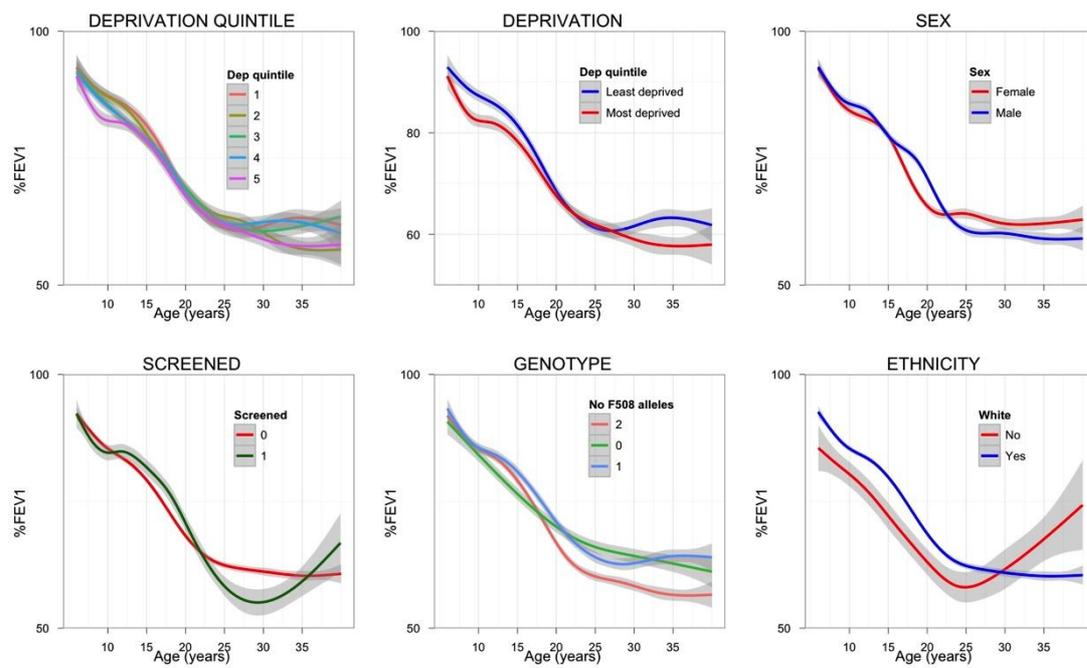
Standard errors in parentheses, birthyear coefficients not shown

The deprivation effect is multiplied by 58 to generate the contrast between the mid point of the least and most deprived quintile.

Age2 is the coefficient for the split line at age 22 years

**%FEV<sub>1</sub>**

**Figure 91: Exploratory analysis showing smoothed means of %FEV<sub>1</sub> score versus age, stratified by covariates**



**Table 26: Final regression models for %FEV<sub>1</sub> in >18 age group**

	<i>baseline</i>	<i>final</i>	<i>final_plus</i>
Constant	66.36305*** (1.92358)	66.98845*** (2.00681)	67.28886*** (2.10252)
Age-18	-0.83027*** (0.12041)	-0.82954*** (0.12040)	-0.48092** (0.18083)
age2	0.17202 (0.17310)	0.17140 (0.17309)	-0.35969 (0.25009)
Number of F508 alleles: 0/2	0.76615 (1.49731)	0.81024 (1.49776)	0.78226 (1.48700)
Number of F508 alleles: 1/2	4.25191*** (1.10186)	4.28061*** (1.10210)	4.14929*** (1.08994)
Male	7.33010*** (1.02742)	7.32078*** (1.02738)	7.18506*** (1.01269)
Non-white	-5.03984 (3.08501)	-4.78147 (3.09386)	-4.64612 (3.05171)
Screened	2.92309 (1.72673)	2.91567 (1.72665)	2.81094 (1.70461)
Age-18 x Number of F508 alleles: 0/2	0.52820* (0.21359)	0.52810* (0.21357)	0.46467* (0.21391)
Age-18 x Number of F508 alleles: 1/2	-0.03008 (0.14545)	-0.03021 (0.14543)	-0.07093 (0.14551)
Age-18 x male	-1.44383*** (0.13344)	-1.44444*** (0.13343)	-1.44413*** (0.13260)
Age-18 x nonwhite	0.22305 (0.42572)	0.21989 (0.42569)	0.20467 (0.42259)
Age-18 x screened	-0.54162* (0.24882)	-0.54524* (0.24882)	-0.51555* (0.24715)
age2 x Number of F508 alleles: 0/2	-0.70699* (0.30242)	-0.70463* (0.30240)	-0.58411 (0.30309)
age2 x Number of F508 alleles: 1/2	0.09007 (0.20501)	0.09044 (0.20500)	0.17869 (0.20603)
age2 x male	1.83791*** (0.19072)	1.83810*** (0.19071)	1.85091*** (0.18987)
age2 x nonwhite	-0.11155 (0.68181)	-0.11334 (0.68174)	-0.03015 (0.67764)
age2 x screened	0.49587 (0.43627)	0.49778 (0.43623)	0.48840 (0.43358)
Deprivation score		-0.02720 (0.02491)	-0.02632 (0.02448)
<i>Pseudomonas</i> colonisation			-1.28927** (0.49353)
CFRD			-3.37204*** (0.82035)
Pancreatic insufficiency			0.74468 (0.77510)
Age-18 x <i>Pseudomonas</i> colonisation			-0.16789 (0.09803)
Age-18 x CFRD			0.32658* (0.14804)
Age-18 x Pancreatic insufficiency			-0.23598 (0.14847)
age2 x <i>Pseudomonas</i> colonisation			0.18743 (0.13566)
age2 x CFRD			-0.09181 (0.19179)
age2 x Pancreatic insufficiency			0.32312 (0.19945)
Log-likelihood	-79717.51480	-79716.91908	-79639.43492
Deviance	159435.02959	159433.83816	159278.86984
AIC	159547.02959	159547.83816	159410.86984
BIC	159989.57167	159998.28278	159932.43729
N	19981	19981	19981
Groups	4026	4026	4026

\*P < 0.05, \*\* P < 0.01, \*\*\* P < 0.001

Standard errors in parentheses, birthyear coefficients not shown

The deprivation effect is multiplied by 58 to generate the contrast between the mid point of the least and most deprived quintile  
age2 is the coefficient for the split line at age 25 years

## *Pseudomonas* colonisation

**Table 27: Final regression models for *P. aeruginosa* colonisation in <18 age group**

	<i>baseline</i>	<i>final</i>	<i>final_plus</i>	<i>final_plus2</i>	<i>final_plus3</i>
Constant	-3.645*** (0.318)	-3.918*** (0.326)	-5.002*** (0.354)	-5.138*** (0.581)	-4.933*** (0.355)
Age	0.216*** (0.038)	0.218*** (0.038)	0.230*** (0.038)	0.352*** (0.083)	0.231*** (0.038)
Age^2	-0.005** (0.002)	-0.005** (0.002)	-0.005*** (0.002)	-0.009** (0.003)	-0.006*** (0.002)
Number of F508 alleles: 0/2	-0.500** (0.179)	-0.520** (0.179)	-0.328 (0.179)	-0.700*** (0.200)	-0.351* (0.179)
Number of F508 alleles: 1/2	-0.468*** (0.115)	-0.478*** (0.115)	-0.353** (0.114)	-0.550*** (0.128)	-0.341** (0.114)
Male	-0.334** (0.106)	-0.338** (0.106)	-0.335** (0.105)	-0.326** (0.118)	-0.328** (0.105)
Deprivation score		0.011*** (0.003)	0.011*** (0.003)	0.012*** (0.004)	0.012*** (0.003)
CFRD			0.588*** (0.147)		0.584*** (0.147)
Pancreatic insufficiency			1.027*** (0.134)		1.030*** (0.134)
%FEV1				-0.032*** (0.002)	
Screened					-0.576*** (0.150)
Log-likelihood	-10119.281	-10112.984	-10072.582	-7783.267	-10064.161
Deviance	20238.562	20225.968	20145.163	15566.534	20128.322
AIC	20316.562	20305.968	20229.163	15636.534	20214.322
BIC	20635.959	20633.556	20573.130	15909.659	20566.478
N	26627	26627	26627	18098	26627
Groups	5648	5648	5648	4334	5648

\*P < 0.05, \*\* P < 0.01, \*\*\* P < 0.001

NB parameters represent log-odds

Standard errors in parentheses, birthyear coefficients not shown

The deprivation effect is multiplied by 58 to generate the contrast between the mid point of the least and most deprived quintile

**Table 28: Final regression models for *P. aeruginosa* colonisation in >18 age group.**

	<i>baseline</i>	final	final plus1	final plus2	final plus3
Constant	1.482*** (0.217)	1.274*** (0.227)	-1.273*** (0.245)	-1.297*** (0.245)	1.307*** (0.221)
Age-18	0.005 (0.010)	0.005 (0.010)	0.022* (0.011)	0.022* (0.011)	-0.022* (0.010)
Number of F508 alleles: 0/2	-1.110*** (0.150)	-1.138*** (0.150)	-0.640*** (0.151)	-0.637*** (0.151)	-0.987*** (0.146)
Number of F508 alleles: 1/2	-0.713*** (0.111)	-0.723*** (0.111)	-0.273* (0.110)	-0.275* (0.110)	-0.534*** (0.107)
Male	-0.188 (0.102)	-0.182 (0.102)	-0.217* (0.101)	-0.221* (0.101)	-0.168 (0.099)
Deprivation score		0.010** (0.003)	0.012*** (0.003)	0.012*** (0.003)	0.008** (0.003)
CFRD			0.703*** (0.086)	0.707*** (0.086)	
Pancreatic insufficiency			2.389*** (0.097)	2.385*** (0.097)	
Screened				0.394 (0.204)	
%FEV <sub>1</sub>					-0.040*** (0.002)
Log-likelihood	-10017.703	-10013.739	-9657.853	-9656.182	-9344.819
Deviance	20035.407	20027.477	19315.706	19312.364	18689.638
AIC	20115.407	20109.477	19401.706	19400.364	18773.638
BIC	20430.169	20432.108	19740.075	19746.602	19102.732
N	19323	19323	19323	19323	18687
Groups	4020	4020	4020	4020	3981

\*P < 0.05, \*\* P < 0.01, \*\*\* P < 0.001

NB parameters represent log-odds

Standard errors in parentheses, birthyear coefficients not shown

The deprivation effect is multiplied by 58 to generate the contrast between the mid point of the least and most deprived quintile.

## Any IV antibiotic therapy

The presentation of the model tables for treatment outcomes follows the same basic format. The severity model is presented first, adjusted for baseline covariates, and disease severity. Then deprivation is added (final model – the main results are from these models), and then other covariates of potential interest are added (final\_plus).

**Table 29: Final regression models for any IV therapy age 18 to <40**

	<i>severity</i>	<i>final</i>	<i>final_plus</i>
Constant	-1.541*** (0.162)	0.587** (0.186)	0.515** (0.188)
Age-18	0.122*** (0.018)	-0.011 (0.019)	-0.009 (0.019)
(Age-18)^2	-0.004*** (0.001)	-0.001 (0.001)	-0.001 (0.001)
Number of F508 alleles: 0/2	0.016 (0.106)	-0.527*** (0.114)	-0.553*** (0.116)
Number of F508 alleles: 1/2	-0.153* (0.077)	-0.341*** (0.083)	-0.326*** (0.084)
Male	-0.305*** (0.070)	-0.743*** (0.077)	-0.752*** (0.078)
%FEV <sub>1</sub>	-0.034*** (0.001)	-0.043*** (0.001)	-0.041*** (0.002)
<i>Pseudomonas</i> colonisation	1.053*** (0.054)	1.617*** (0.055)	1.599*** (0.056)
Deprivation		0.011*** (0.002)	0.011*** (0.003)
BMI SD score			-0.142*** (0.029)
Log-likelihood	-9721.856	-9040.389	-8948.984
Deviance	19443.712	18080.777	17897.969
AIC	19529.712	18168.777	17987.969
BIC	19866.439	18513.335	18339.961
N	18599	18599	18436
Groups	3971	3971	3962

\*P < 0.05, \*\* P < 0.01, \*\*\* P < 0.001

NB parameters represent log(IV days)

Standard errors in parentheses, birthyear coefficients not shown

The deprivation effect is multiplied by 58 to generate the contrast between the mid point of the least and most deprived quintile.

**Table 30: Robustness test: regression models for use of any IV therapy 5 to <18**

	final	Deprivation z score	Data <2000 excluded	Cepacia added	Care centre
Constant	-1.800*** (0.202)	-1.459*** (0.195)	-1.812*** (0.205)	-1.803*** (0.202)	-1.035*** (0.270)
Age-5	0.340*** (0.036)	0.340*** (0.036)	0.336*** (0.037)	0.339*** (0.036)	0.213*** (0.038)
(Age-5)^2	-0.016*** (0.002)	-0.016*** (0.002)	-0.016*** (0.003)	-0.016*** (0.002)	-0.006* (0.003)
Number of F508 alleles: 0/2	-0.533*** (0.127)	-0.533*** (0.127)	-0.509*** (0.128)	-0.526*** (0.127)	-0.243* (0.119)
Number of F508 alleles: 1/2	-0.388*** (0.079)	-0.388*** (0.079)	-0.385*** (0.080)	-0.389*** (0.079)	-0.294*** (0.074)
Male	-0.374*** (0.073)	-0.374*** (0.073)	-0.352*** (0.074)	-0.378*** (0.073)	-0.400*** (0.068)
fev_cent	-0.036*** (0.002)	-0.036*** (0.002)	-0.036*** (0.002)	-0.035*** (0.002)	-0.040*** (0.002)
<i>Pseudomonas</i> colonisation	1.720*** (0.065)	1.720*** (0.065)	1.684*** (0.066)	1.725*** (0.065)	1.842*** (0.065)
Deprivation	0.016*** (0.002)		0.016*** (0.002)	0.016*** (0.002)	0.010*** (0.002)
Deprivation Z score		0.247*** (0.036)			
<i>B.Cepacia</i>				1.095*** (0.240)	
Log-likelihood	-9447.085	-9447.356	-8978.788	-9436.327	-8996.254
Deviance	18894.170	18894.711	17957.576	18872.655	17992.507
AIC	18966.170	18966.711	18023.576	18946.655	18322.507
BIC	19246.877	19247.418	18279.144	19235.159	19609.079
N	17987	17987	17060	17987	17987
Groups	4321	4321	4220	4321	4321

\*P < 0.05, \*\* P < 0.01, \*\*\* P < 0.001

NB parameters represent log-odds

Standard errors in parentheses, birthyear coefficients not shown

The deprivation effect is multiplied by 58 to generate the contrast between the mid point of the least and most deprived quintile.

The deprivation effect in Z scores is multiplied by 3.56 to generate the contrast between the mid point of the least and most deprived quintile

### Any home IV therapy

**Table 31: Final regression models for any home IV therapy age 5-18**

	severity	final
Constant	-3.705*** (0.247)	-4.002*** (0.296)
Age-5	0.187*** (0.014)	0.247*** (0.022)
Number of F508 alleles: 0/2	-0.882*** (0.190)	-0.836*** (0.192)
Number of F508 alleles: 1/2	-0.282* (0.115)	-0.256* (0.116)
Male	-0.291** (0.107)	-0.308** (0.108)
%FEV <sub>1</sub>	-0.022*** (0.002)	-0.022*** (0.002)
<i>Pseudomonas</i> colonisation	1.190*** (0.074)	1.203*** (0.074)
Deprivation		0.012 (0.007)
Age-5 x deprivation		-0.003*** (0.001)
Log-likelihood	-7773.296	-7767.323
Deviance	15546.592	15534.646
AIC	15614.592	15606.646
BIC	15879.704	15887.353
N	17987	17987
Groups	4321	4321

\*P < 0.05, \*\* P < 0.01, \*\*\* P < 0.001

NB parameters represent log-odds

Standard errors in parentheses, birthyear coefficients not shown

The deprivation effect is multiplied by 58 to generate the contrast between the mid point of the least and most deprived quintile.

**Table 32: Final regression models for any home IV therapy age 18-40**

	<i>severity</i>	<i>final</i>
Constant	-0.762*** (0.186)	-0.770*** (0.186)
Age-18	0.034 (0.020)	0.034 (0.020)
(Age-18)^2	-0.003** (0.001)	-0.003** (0.001)
Number of F508 alleles: 0/2	-0.894*** (0.125)	-0.869*** (0.125)
Number of F508 alleles: 1/2	-0.245** (0.088)	-0.234** (0.088)
Male	-0.972*** (0.081)	-0.978*** (0.081)
%FEV <sub>1</sub>	-0.031*** (0.001)	-0.031*** (0.001)
<i>Pseudomonas</i> colonisation	1.195*** (0.057)	1.199*** (0.057)
Deprivation		-0.009*** (0.003)
Log-likelihood	-9617.911	-9612.461
Deviance	19235.821	19224.922
AIC	19321.821	19312.922
BIC	19658.548	19657.480
N	18599	18599
Groups	3971	3971

\*P < 0.05, \*\* P < 0.01, \*\*\* P < 0.001

NB parameters represent log-odds

Standard errors in parentheses, birthyear coefficients not shown

The deprivation effect is multiplied by 58 to generate the contrast between the mid point of the least and most deprived quintile.

**Table 33: Final regression models for amount of any IV therapy age 5 to <18**

	<i>severity</i>	<i>final</i>	<i>final_plus</i>
Constant	2.962*** (0.045)	2.900*** (0.048)	2.901*** (0.048)
Age-5	0.031*** (0.003)	0.031*** (0.003)	0.031*** (0.003)
Number of F508 alleles: 0/2	-0.008 (0.036)	-0.013 (0.036)	-0.013 (0.036)
Number of F508 alleles: 1/2	-0.041 (0.022)	-0.043* (0.022)	-0.043* (0.022)
Male	-0.091*** (0.020)	-0.092*** (0.020)	-0.092*** (0.020)
%FEV <sub>1</sub>	-0.008*** (0.000)	-0.008*** (0.000)	-0.008*** (0.000)
<i>Pseudomonas</i> colonisation	0.224*** (0.015)	0.222*** (0.015)	0.223*** (0.015)
Deprivation		0.003*** (0.001)	0.003*** (0.001)
BMI SD score			-0.000 (0.002)
Log-likelihood	-7924.630	-7915.716	-7905.456
Deviance	15849.260	15831.432	15810.912
AIC	15919.260	15903.432	15884.912
BIC	16166.953	16158.201	16146.712
N	8751	8751	8740
Groups	3004	3004	3003

\*P < 0.05, \*\* P < 0.01, \*\*\* P < 0.001

NB parameters represent log(IV days)

Standard errors in parentheses, birthyear coefficients not shown

The deprivation effect is multiplied by 58 to generate the contrast between the mid point of the least and most deprived quintile.

**Table 34: Final regression models for amount of IV therapy age 18 to <40**

	<i>severity</i>	<i>final</i>	<i>final_plus</i>
Constant	3.234*** (0.044)	3.198*** (0.046)	3.174*** (0.046)
Age-18	-0.003 (0.002)	-0.003 (0.002)	-0.002 (0.002)
Number of F508 alleles: 0/2	-0.095** (0.033)	-0.098** (0.033)	-0.099** (0.033)
Number of F508 alleles: 1/2	-0.020 (0.023)	-0.022 (0.023)	-0.019 (0.023)
Male	-0.157*** (0.021)	-0.156*** (0.021)	-0.159*** (0.021)
%FEV <sub>1</sub>	-0.012*** (0.000)	-0.012*** (0.000)	-0.011*** (0.000)
<i>Pseudomonas</i> colonisation	0.192*** (0.016)	0.191*** (0.016)	0.191*** (0.016)
Deprivation		0.002** (0.001)	0.002* (0.001)
BMI SD score			-0.032*** (0.008)
Log-likelihood	-10632.661	-10629.278	-10538.959
Deviance	21265.322	21258.555	21077.918
AIC	21349.322	21344.555	21165.918
BIC	21656.068	21658.604	21486.925
N	10976	10976	10890
Groups	3052	3052	3046

\*P < 0.05, \*\* P < 0.01, \*\*\* P < 0.001

NB parameters represent log(IV days)

Standard errors in parentheses, birthyear coefficients not shown

The deprivation effect is multiplied by 58 to generate the contrast between the mid point of the least and most deprived quintile.

## Nutritional therapy

**Table 35: Final regression models for any nutritional support <18 age**

	<i>severity</i>	<i>final</i>	<i>final_plus</i>
Constant	-4.391*** (0.229)	-4.607*** (0.235)	-5.192*** (0.482)
Age	0.514*** (0.029)	0.511*** (0.029)	0.566*** (0.070)
Age <sup>2</sup>	-0.022*** (0.001)	-0.022*** (0.001)	-0.024*** (0.003)
Number of F508 alleles: 0/2	-0.375** (0.126)	-0.396** (0.126)	-0.544*** (0.148)
Number of F508 alleles: 1/2	-0.412*** (0.081)	-0.419*** (0.081)	-0.392*** (0.093)
Male	0.125 (0.075)	0.120 (0.074)	0.207* (0.086)
BMI SD score	-0.740*** (0.029)	-0.735*** (0.029)	-0.826*** (0.038)
<i>Pseudomonas</i> colonisation	1.019*** (0.055)	1.012*** (0.055)	1.081*** (0.063)
Deprivation		0.010*** (0.002)	0.010*** (0.003)
%FEV <sub>1</sub>			-0.014*** (0.002)
Log-likelihood	-13462.592	-13452.527	-9645.127
Deviance	26925.184	26905.054	19290.255
AIC	27007.184	26989.054	19364.255
BIC	27346.513	27336.660	19656.601
N	29037	29037	19955
Groups	5754	5754	4442

\*P < 0.05, \*\* P < 0.01, \*\*\* P < 0.001

NB parameters represent log-odds

Standard errors in parentheses, birthyear coefficients not shown

The deprivation effect is multiplied by 58 to generate the contrast between the mid point of the least and most deprived quintile.

**Table 36. Final regression models for any nutritional support age 18-40**

	<i>severity</i>	<i>final</i>	<i>final_plus</i>
Constant	-3.232*** (0.193)	-3.563*** (0.208)	-3.836*** (0.230)
Age-18	0.133*** (0.021)	0.160*** (0.023)	0.200*** (0.024)
(Age-18)^2	-0.005*** (0.001)	-0.005*** (0.001)	-0.007*** (0.001)
Number of F508 alleles: 0/2	-0.543*** (0.131)	-0.442*** (0.133)	-0.487*** (0.145)
Number of F508 alleles: 1/2	-0.309** (0.094)	-0.306** (0.096)	-0.319** (0.105)
Male	0.409*** (0.087)	0.388*** (0.088)	0.400*** (0.096)
BMI SD score	-0.882*** (0.030)	-0.853*** (0.031)	-0.711*** (0.035)
<i>Pseudomonas</i> colonisation	1.155*** (0.058)	1.308*** (0.059)	1.283*** (0.063)
Deprivation		0.015*** (0.003)	0.016*** (0.003)
%FEV <sub>1</sub>			-0.021*** (0.002)
Log-likelihood	-9250.602	-9210.385	-8938.167
Deviance	18501.205	18420.770	17876.333
AIC	18591.205	18512.770	17970.333
BIC	18947.097	18876.571	18341.341
N	20105	20105	19807
Groups	4040	4040	4017

\*P < 0.05, \*\* P < 0.01, \*\*\* P < 0.001

NB parameters represent log-odds

Standard errors in parentheses, birthyear coefficients not shown

The deprivation effect is multiplied by 58 to generate the contrast between the mid point of the least and most deprived quintile.

## DNase

**Table 37. Final regression models for any DNase <18 age**

	<i>severity</i>	<i>final</i>	<i>final_plus</i>
Age	3.489*** (0.109)	3.528*** (0.111)	3.642*** (0.148)
Age^2	-0.105*** (0.004)	-0.106*** (0.004)	-0.111*** (0.006)
Number of F508 alleles: 0/2	-0.351 (0.331)	-0.427 (0.343)	-0.368 (0.276)
Number of F508 alleles: 1/2	-0.758*** (0.226)	-0.751** (0.234)	-0.603*** (0.182)
Male	-0.722*** (0.211)	-0.670** (0.219)	-0.567*** (0.169)
<i>Pseudomonas</i> colonisation	0.935*** (0.100)	0.959*** (0.101)	0.918*** (0.100)
BMI SD score	-0.217*** (0.060)	-0.315*** (0.062)	-0.010 (0.024)
Deprivation		-0.011 (0.007)	-0.016** (0.005)
%FEV <sub>1</sub>			-0.052*** (0.003)
Log-likelihood	-8936.299	-8935.158	-7496.707
Deviance	17872.598	17870.316	14993.414
AIC	17954.598	17954.316	15067.414
BIC	18293.940	18301.935	15359.765
N	29046	29046	19958
Groups	5757	5757	4443

\*P < 0.05, \*\* P < 0.01, \*\*\* P < 0.001

NB parameters represent log-odds

Standard errors in parentheses, birthyear coefficients not shown

The deprivation effect is multiplied by 58 to generate the contrast between the mid point of the least and most deprived quintile.

**Table 38: Final regression models for any DNase age 18-40**

	<i>severity</i>	<i>final</i>
Constant	-2.200*** (0.225)	-1.845*** (0.234)
Age-18	0.148*** (0.009)	0.148*** (0.009)
Number of F508 alleles: 0/2	-0.775*** (0.156)	-0.719*** (0.155)
Number of F508 alleles: 1/2	-0.261* (0.111)	-0.239* (0.111)
Male	-0.289** (0.103)	-0.295** (0.103)
<i>Pseudomonas</i> colonisation	1.292*** (0.062)	1.296*** (0.062)
%FEV <sub>1</sub>	-0.046*** (0.002)	-0.046*** (0.002)
Deprivation		-0.017*** (0.003)
Log-likelihood	-8932.258	-8919.771
Deviance	17864.517	17839.541
AIC	17948.517	17925.541
BIC	18280.423	18265.351
N	19981	19981
Groups	4026	4026

\*P < 0.05, \*\* P < 0.01, \*\*\* P < 0.001

NB parameters represent log-odds

Standard errors in parentheses, birthyear coefficients not shown

The deprivation effect is multiplied by 58 to generate the contrast between the mid point of the least and most deprived quintile.

## Any inhaled therapy <18

**Table 39: Final regression models for any inhaled therapy <18 age**

	<i>severity</i>	<i>final</i>
Constant	-0.265 (0.379)	-0.100 (0.391)
Age	0.279*** (0.055)	0.262*** (0.055)
Age^2	-0.007** (0.002)	-0.006** (0.002)
Number of F508 alleles: 0/2	-0.516*** (0.144)	-0.642*** (0.144)
Number of F508 alleles: 1/2	-0.342*** (0.091)	-0.544*** (0.091)
Male	-0.006 (0.085)	-0.020 (0.085)
<i>Pseudomonas</i> colonisation	2.126*** (0.076)	2.379*** (0.079)
%FEV <sub>1</sub>	-0.020*** (0.002)	-0.019*** (0.002)
Deprivation		-0.007** (0.003)
Log-likelihood	-9623.226	-9573.725
Deviance	19246.452	19147.451
AIC	19316.452	19219.451
BIC	19593.039	19503.940
N	19980	19980
Groups	4445	4445

\*P < 0.05, \*\* P < 0.01, \*\*\* P < 0.001

NB parameters represent log-odds

Standard errors in parentheses, birthyear coefficients not shown

The deprivation effect is multiplied by 58 to generate the contrast between the mid point of the least and most deprived quintile.

**Table 40: Final regression models for any inhaled therapy 18-40 years**

	<i>severity</i>	<i>final</i>
Constant	1.889*** (0.547)	2.204*** (0.550)
ageatvisityrs	-0.148*** (0.040)	-0.147*** (0.041)
I(ageatvisityrs^2)	0.003*** (0.001)	0.003*** (0.001)
Number of F508 alleles: 0/2	-0.505*** (0.076)	-0.468*** (0.076)
Number of F508 alleles: 1/2	-0.240*** (0.055)	-0.225*** (0.055)
Male	-0.101* (0.050)	-0.111* (0.050)
<i>Pseudomonas</i> colonisation	1.391*** (0.041)	1.419*** (0.042)
%FEV <sub>1</sub>	-0.011*** (0.001)	-0.012*** (0.001)
Deprivation		-0.016*** (0.002)
Log-likelihood	-10255.477	-10198.083
Deviance	20510.954	20396.166
AIC	20600.954	20488.166
BIC	20956.568	20851.683
N	19981	19981
Groups	4026	4026

\*P < 0.05, \*\* P < 0.01, \*\*\* P < 0.001

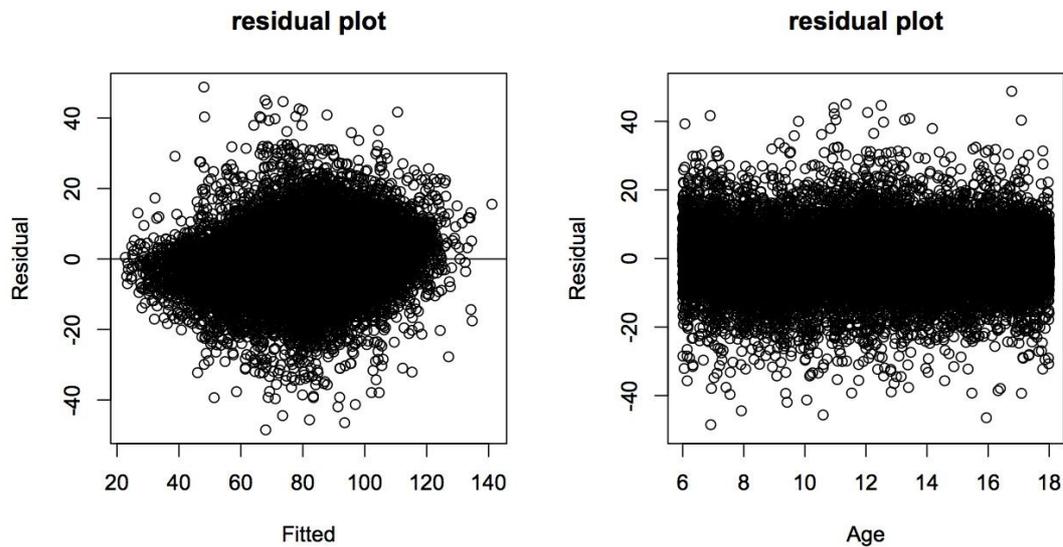
NB parameters represent log-odds

Standard errors in parentheses, birthyear coefficients not shown

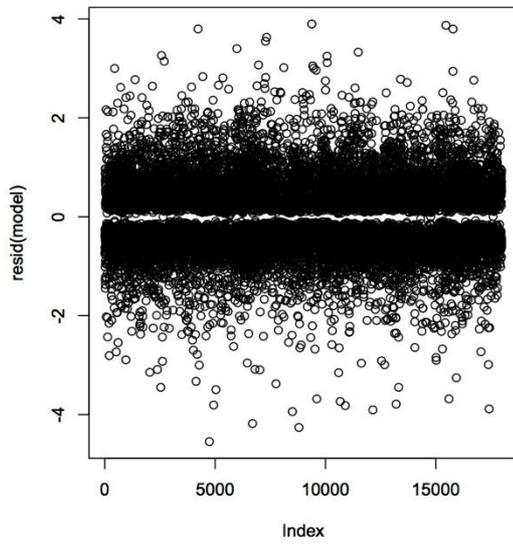
The deprivation effect is multiplied by 58 to generate the contrast between the mid point of the least and most deprived quintile.

## Illustrative regression diagnostics

**Figure 92: %FEV<sub>1</sub> < 18 years, final model**



**Figure 93: Any IV therapy < 18 years, final model**



## Appendix 5 (Pertaining to Chapter 5)

**Table 41: Log odds for the final GLMMs, with added educational variable.**

	<i>Baseline + Severity + Time in hospital + Deprivation*%FEV<sub>1</sub></i>	<i>Plus education variable (NB models not nested)</i>
Constant	2.05821*** (0.17127)	-0.73603 (0.43094)
Deprivation quintile 2/1	-0.27946 (0.20153)	-0.21829 (0.23348)
Deprivation quintile 3/1	-0.97564*** (0.19987)	-0.70394** (0.24049)
Deprivation quintile 4/1	-1.42676*** (0.19752)	-0.99734*** (0.23896)
Deprivation quintile 5/1	-2.66327*** (0.20678)	-1.98129*** (0.25776)
%FEV <sub>1</sub>	0.01257** (0.00487)	0.01111 (0.00579)
Hospital IV days	-0.02309*** (0.00209)	-0.02226*** (0.00282)
BMI SDS	0.10406* (0.04304)	
age	0.14598*** (0.01616)	0.21636*** (0.02064)
age^2	-0.02338*** (0.00164)	-0.02087*** (0.00212)
Birthyear	-0.03949** (0.01427)	-0.03994* (0.01826)
Male/Female	0.40087** (0.12499)	0.45083** (0.15167)
Deprivation quintile 2/1 x %FEV <sub>1</sub>	0.01031 (0.00661)	0.00613 (0.00797)
Deprivation quintile 3/1 x %FEV <sub>1</sub>	0.00979 (0.00641)	0.00676 (0.00819)
Deprivation quintile 4/1 x %FEV <sub>1</sub>	0.00581 (0.00638)	0.01120 (0.00796)
Deprivation quintile 5/1 x %FEV <sub>1</sub>	0.01642* (0.00695)	0.01814* (0.00881)
Ranef - id	6.92797 (2.63210)	6.05039 (2.45980)
Ranef	0.19489 (0.44146)	0.18887 (0.43460)
Highest education: 2/1 (1 = "less than high school")		2.08872*** (0.40422)
Highest education: 3/1		3.09579*** (0.44268)
Highest education: 4/1		3.21147*** (0.41106)
Highest education: 5/1		2.76440*** (0.39619)
Log-likelihood	-7545.11112	-4606.94109
Deviance	15090.22224	9213.88218
AIC	15128.22224	9257.88218
BIC	15273.45955	9415.17414
N	15430	9411
Groups	3451	2008

## Appendix 6 (Pertaining to Chapter 6)

### Statistical Package (R) Programming Code

```
#generating spaghetti plots for 10 individuals

ns <- 10
samp <- sample(d$Sid,ns)
sampd <- subset(d, d$id %in% samp)
ids <- unique(sampd$Sid)

plot(sampd$age[sampd$Sid == ids[1]],sampd$fev1[sampd$Sid == ids[1]], type="l", ylim=c(0,150),xlab="Age",main="REAL
DATA",ylab="% Predicted FEV1")
for (i in 2:ns)
  lines(sampd$age[sampd$Sid == ids[i]],sampd$fev1[sampd$Sid == ids[i]],col=i)

#generating variogram

variogram<-function(id,time,residual,u.max=NULL,u.increment=1) {
#
# variogram function adapted from the geoR library, to deal with
# longitudinal data-sets with long individual time series, geoR
# library must be installed before use.
#
# Arguments:
# id: identifier for individual subjects
# time: time at which measurement is made
# residual: corresponding residual from model for mean response profiles
# u.max: maximum time-separation at which variogram is estimated (optional, but recommended)
# u.increment: increment between successive time-separations (not used when u.max=NULL)
#
# Result: a list with components:
# u: time-separations at which variogram is estimated
# v: corresponding variogram estimates
# n: number of pairs contributing to each variogram estimate
# sigmasquared: variogram-based estimate of process variance (sum of between-subject
# and within-subject components)
#
# NOTES: 1. when data include replicated measurements at a common time within one or more subjects,
# the geoR library generates warning messages...these can safely be ignored in the
# present context
# 2. when data include subjects with only one non-missing response, variogram calculation will fail
#
nid<-length(id)
nt<-length(time)
nr<-length(residual)
check1<-c(nid-nt,nid-nr)
if (max(abs(check1))>0) print("Bad data: unequal lengths amongst id, time and response")
check2<-table(id)
if (min(check2)<2) print("Bad data: at least one subject with only 1 response")
library(geoR)
uid<-unique(id)
nid<-length(uid)
if (length(u.max)==0) {
  u1<-min(time)
  u2<-max(time)
  h<-(u2-u1)/40
  u<-((1:20)-0.5)*u.increment
} else {
  u<-u.increment*((1:round(u.max/u.increment))-0.5)
}
nu<-length(u)
u.all<-NULL
v.all<-NULL
n.within<-rep(NA,nid)
mean.within<-rep(NA,nid)
var.within<-rep(NA,nid)
for (i in 1:length(uid)) {
  take<-(id==uid[i])
  if (sum(take)>1) {
    x<-time[take]
    y<-rep(0,length(x))
```

```

z<-residual[take]
xyz<-as.geodata(cbind(x,y,z))
vario<-variog(xyz,option="cloud",messages=F)
u.all<-c(u.all,vario$u)
v.all<-c(v.all,vario$v)
n.within[i]<-length(!is.na(z))
mean.within[i]<-mean(z[!is.na(z)])
var.within[i]<-var(z[!is.na(z)])
}
}
sigmasquared<-sum(var.within[n.within>=2]*(n.within[n.within>=2]-1))/sum(n.within[n.within>=2]-1) +
var(mean.within[!is.na(mean.within)])
nugget<-min(u.all)==0
u.breaks<-c(0,u+0.5*u.increment)
if (nugget==TRUE) {
u.breaks<-c(0,u.breaks)
u<-c(0,u)
}
nu<-length(u)
v<-rep(NA,nu)
n<-rep(0,nu)
if (nugget==T) {
take<-u.all==0
n[1]<-sum(take)
v[1]<-mean(v.all[take],na.rm=T)
} else {
take<-(u.all>=u.breaks[1])&(u.all<u.breaks[2])
n[1]<-sum(take)
v[1]<-mean(v.all[take],na.rm=T)
}
for (i in 2:nu) {
take<-(u.all>=u.breaks[i])&(u.all<u.breaks[i+1])
n[i]<-sum(take)
v[i]<-mean(v.all[take],na.rm=T)
}
list(u=u,v=v,n=n,sigmasquared=sigmasquared)
}

# Functions required for exploratory analysis

average.by.age<-function(x,y,x.lowest,x.increment,x.highest) {
nbreaks<-ceiling((range(x)[2]-range(x)[1])/x.increment)
breaks<-x.lowest+x.increment*(0:nbreaks)
yvec<-rep(0,nbreaks)
for (i in 1:nbreaks) {
take<-(x>breaks[i])&(x<=breaks[i+1])
yvec[i]<-mean(y[take],na.rm=T)
}
xvec<-(breaks[2:(nbreaks+1)]+breaks[1:nbreaks])/2
list(x=xvec,y=yvec)
}

data <- d
smooth.trend<-average.by.age(data$age,data$fev1,5,1,60)
plot(data$age,data$fev1,pch=".",xlab="age (years)",ylab="%FEV1")
lines(smooth.trend$x,smooth.trend$y,type="l",lwd=2,col="red")

# add columns to data for smoothed response and residuals

yfit<-smooth.trend$y
xfit<-smooth.trend$x
N<-dim(data)[1]
ysmooth<-rep(NA,N)
for (i in 1:N) {
take<-floor(data$age[i])==floor(xfit)
if (sum(take)>0) ysmooth[i]<-yfit[take]
}
data$smooth<-ysmooth
data$res<-data$fev1-data$smooth

d$res <- data$res
x<-d$age
z<-d$res
id<-d$id
data.v<- variogram(id,x,z,u.max=30)

```

#plotting theoretical variogram trace (red line in **Figure 70**)

```
x <- seq(0,30,0.1)
torsq <- 0.1500497* 16.41201^2
sigsq <- (1-0.1500497)* 16.41201^2
nusq <- 18.87013^2
phi <- 6.7211035
y5 <- torsq + sigsq*(1-exp(-x/phi))
lines(x,y5,col="red",lwd=2)
lines(x,rep(torsq+sigsq+nusq,length(x)),col="red",lwd=2,lty=2)
```

#final multivariate model specification – this runs overnight on this dataset on a 2.8GHz Intel Core 2 Mac Book Pro (8GB memory), running 64-bit version of R

```
library(nlme)
expl <- corExp(value=c(7,(40/300)),form=~age2|id,nugget=T)
expl <- Initialize(expl,d)
m15co <- lme(fev1 ~ sex+DM2+ age2*(cohort + Plb) +age3, data=d, random= ~ 1|id, method="ML",correlation=expl)
summary(m15co)
intervals(m15co)
```

# simulation of data from a class of longitudinal models, irregularly

# observed in time

#

```
mu<-function(t,theta) {
```

#

# Arguments

# t: time (vector of non-negative real numbers)

# theta: vector of parameters that define the mean response as a function of time

#

# Result

# Vector giving the values of the mean response at times corresponding to

# each element of the vector t

#

# Comment

# The following code is indicative only: it defines the mean response as

# linear in time, with intercept theta[1] and slope theta[2] but can be

# replaced by any other code that operates on vectors t and theta supplied

# as arguments

#

```
theta[1]+theta[2]*t
```

```
}
```

#

```
vmat<-function(t,nusqA,nusqB,rhoAB,tausq,sigmasq,phi,kappa=0.5) {
```

#

# Arguments

# t: time (vector of non-negative real numbers)

# nusqA: variance of random intercept

# nusqB: variance of random slope

# rhoAB: correlation between random intercept and slope

# tausq: measurement error variance

# sigmasq: variance of serially correlated component

# phi: scale parameter of Matern correlation function

# kappa: shape parameter of Matern correlation function (defaults to exponential)

#

# Result

# Variance matrix of sequence of measurements at times corresponding to

# each element of the vector t

```
nt<-length(t)
```

```
vAB<-matrix(c(nusqA,rep(rhoAB*sqrt(nusqA*nusqB),2),nusqB),2,2)
```

```
xmat<-cbind(rep(1,nt),t)
```

```
V1<-xmat%*%vAB%*%t(xmat)
```

```
V2<-sigmasq*matern(abs(outer(t,t,"-")),phi,kappa)
```

```
V3<-tausq*diag(rep(1,nt))
```

```
V1+V2+V3
```

```
}
```

```
simulate<-function(t,theta,nusqA,nusqB,rhoAB,tausq,sigmasq,phi,kappa=0.5) {
```

#

# Arguments

#

# t: time (vector of non-negative real numbers)

# theta: vector of parameters that define the mean response as a function of time

# nusqA: variance of random intercept

# nusqB: variance of random slope

```

# rhoAB: correlation between random intercept and slope
# tausq: measurement error variance
# sigmasq: variance of serially correlated component
# phi: scale parameter of Matern correlation function
# kappa: shape parameter of Matern correlation function (defaults to exponential)
#
# Result
# Vector containing simulated realisation for a single subject
#
mean.vector<-mu(t,theta)
var.matrix<-vmat(t,nusqA,nusqB,rhoAB,tausq,sigmasq,phi,kappa)
rmvnorm(1,mean.vector,var.matrix)
}

simulate.balanced<-function(ns,t,theta,nusqA,nusqB,rhoAB,tausq,sigmasq,phi,kappa=0.5) {
#
# Arguments
#
# ns: number of subjects
# t: time (vector of non-negative real numbers), common to all subjects
# theta: vector of parameters that define the mean response as a function of time
# nusqA: variance of random intercept
# nusqB: variance of random slope
# rhoAB: correlation between random intercept and slope
# tausq: measurement error variance
# sigmasq: variance of serially correlated component
# phi: scale parameter of Matern correlation function
# kappa: shape parameter of Matern correlation function (defaults to exponential)
#
# Result
# Matrix containing simulated realisations for ns subjects (rows) at
# nt times (columns)
#
mean.vector<-mu(t,theta)
var.matrix<-vmat(t,nusqA,nusqB,rhoAB,tausq,sigmasq,phi,kappa)
rmvnorm(ns,mean.vector,var.matrix)
}

simulate.unbalanced<-function(ns,tlist,theta,nusqA,nusqB,rhoAB,tausq,sigmasq,phi,kappa=0.5) {
#
# Arguments
#
# ns: number of subjects
# tlist: times (list of vectors of non-negative real numbers) unique to each subject
# theta: vector of parameters that define the mean response as a function of time
# nusqA: variance of random intercept
# nusqB: variance of random slope
# rhoAB: correlation between random intercept and slope
# tausq: measurement error variance
# sigmasq: variance of serially correlated component
# phi: scale parameter of Matern correlation function
# kappa: shape parameter of Matern correlation function (defaults to exponential)
#
# Result
# List containing simulated realisation for each subject
#
result<-as.list(1:ns)
for (i in 1:ns) {
mean.vector<-mu(tlist[[i]],theta)
var.matrix<-vmat(tlist[[i]],nusqA,nusqB,rhoAB,tausq,sigmasq,phi,kappa)
result[[i]]<-rmvnorm(1,mean.vector,var.matrix)
}
result
}
#

##code to generate realizations in Figure 72

theta<-c((66.02327 + 25.18973), (-0.26051+-0.72041))
nusqA<- 16.33969^2
nusqB<- 0
rhoAB<- 0
tausq<- 0.1664673*15.57728^2
sigmasq<- (1-0.1664673)*15.57728^2
phi<- 5.9045003

```

```

kappa<-0.5

ns<-4

samp <- sample(d$Sid,ns)
sampd <- subset(d, d$Sid %in% samp)
ids <- unique(sampd$Sid)
tlist<-as.list(1:ns)
for (i in 1:ns)
{ tlist[[i]] <- sampd$Age[sampd$Sid==ids[i]] }

x<-simulate.unbalanced(ns,tlist,theta,nusqA,nusqB,rhoAB,tausq,sigmasq,phi,kappa=0.5)

par(mfrow=c(1,2))

plot(tlist[[1]],x[[1]],type="l", ylim=c(0,120),xlab="Age",main="MODEL",ylab="% Predicted FEV1")
for (i in 2:ns)
lines(tlist[[i]],x[[i]],col=i)

plot(sampd$Age[sampd$Sid == ids[1]],sampd$fev1[sampd$Sid == ids[1]], type="l", ylim=c(0,120),xlab="Age",main="REAL
DATA",ylab="% Predicted FEV1")
for (i in 2:ns)
lines(sampd$Age[sampd$Sid == ids[i]],sampd$fev1[sampd$Sid == ids[i]],col=i)

```

## Appendix (Publications from this thesis)

### Publications:

Taylor-Robinson D, Smyth R, Diggle P, Whitehead M. A longitudinal study of the impact of social deprivation and disease severity on employment status in the UK cystic fibrosis population. PLoS ONE 8(8): e73322. doi:10.1371/journal.pone.0073322

Taylor-Robinson D, Smyth R, Diggle P, Whitehead M (2013) The effect of social deprivation on clinical outcomes and the use of treatments in the UK cystic fibrosis population: a longitudinal study. The Lancet Respiratory Medicine vol 1, issue 2 pp 121-128

Taylor-Robinson D, Whitehead M, Diderichsen F, Olesen H, Smyth RL\*, Diggle P\*. \*Joint last author (2012) Understanding the natural progression in %FEV1 decline in patients with cystic fibrosis: A Longitudinal study. Thorax vol 67 pp 860-866

Taylor-Robinson D, Whitehead M, Diderichsen F, Veber Olesen H, Pressler T, Smyth R, Diggle P. (2013) Author's response: understanding the natural progression in %FEV1 decline in patients with cystic fibrosis: a longitudinal study. Thorax vol 68 issue 3 pp 294-5

Taylor-Robinson DC, Schechter MS (2011) Health inequalities and cystic fibrosis. BMJ vol 343:d4818 (Invited editorial)

Taylor-Robinson D, Olesen HV, Pressler T, Thielen K, Diderichsen F, Diggle P, Smyth R, Whitehead M (2012) The effect of socioeconomic status on outcomes in CF. Pediatric Pulmonology vol 47 issue S35 pp 199-200 (Invited symposium NACF 2012)

### Abstracts:

Taylor-Robinson D, Whitehead M, Smyth R, Diggle P, Henderson R, Barrett J (2012) Longitudinal Changes in Lung Function and risk of Death in Cystic Fibrosis: developing a Joint Model for the UK Population. Journal of Cystic Fibrosis vol 11 (s1) pp 7

Taylor-Robinson D, Whitehead M, Diggle P, Smyth R (2012) The Effect of Social Deprivation on Pseudomonas and Staphylococcal Colonisation in the UK Cystic Fibrosis Population. Journal of Cystic Fibrosis vol 11 (s1) pp 91

Taylor-Robinson D, Whitehead M, Olesen HV, Pressler T, Smyth R, Diggle P. (2011) Understanding the natural progression in FEV1 decline in patients with cystic fibrosis. Journal of Cystic Fibrosis vol Supp 1 issue 10 pp 49 (BEST PULMONOLOGY ABSTRACT European CF meeting, 2011)

Taylor-Robinson D, Whitehead M, Olesen HV, Pressler T, Smyth R, Diggle P (2011) The effect of social deprivation on weight in the UK cystic fibrosis population. Journal of Cystic Fibrosis vol Sup 1 issue 10 pp 72

Taylor-Robinson D, Whitehead M, Diggle P, Smyth R (2010) Social deprivation and rate of decline of lung function in the UK cystic fibrosis population. *Pediatric Pulmonology* vol 45 issue S33 pp 449

# The effect of social deprivation on clinical outcomes and the use of treatments in the UK cystic fibrosis population: a longitudinal study



David C Taylor-Robinson, Rosalind L Smyth, Peter J Diggle, Margaret Whitehead

## Summary

**Background** Poorer socioeconomic circumstances have been linked with worse outcomes in cystic fibrosis. We assessed whether a relation exists between social deprivation and individual's clinical and health-care outcomes.

**Methods** We did a longitudinal registry study of the UK cystic fibrosis population younger than 40 years (8055 people with 49337 observations for weight, the most commonly collected outcome, between Jan 1, 1996, and Dec 31, 2009). We assessed data for weight, height, body-mass index, percent predicted forced expiratory volume in 1 s (%FEV<sub>1</sub>), risk of *Pseudomonas aeruginosa* colonisation, and the use of major cystic fibrosis treatment modalities. We used mixed effects models to assess the association between small-area deprivation and clinical and health-care outcomes, adjusting for clinically important covariates. We give continuous outcomes as mean differences, and binary outcomes as odds ratios, comparing extremes of deprivation quintile.

**Findings** Compared with the least deprived areas, children from the most deprived areas weighed less (standard deviation [SD] score  $-0.28$ , 95% CI  $-0.38$  to  $-0.18$ ), were shorter ( $-0.31$ ,  $-0.40$  to  $-0.21$ , and had a lower body-mass index ( $-0.13$ ,  $-0.22$  to  $-0.04$ ), were more likely to have chronic *P aeruginosa* infection (odds ratio 1.89, 95% CI 1.34 to 2.66), and have a lower %FEV<sub>1</sub> ( $-4.12$  percentage points, 95% CI  $-5.01$  to  $-3.19$ ). These inequalities were apparent very early in life and did not widen thereafter. On a population level, after adjustment for disease severity, children in the most deprived quintile were more likely to receive intravenous antibiotics (odds ratio 2.52, 95% CI 1.92 to 3.17) and nutritional treatments (1.78, 1.44 to 2.20) compared with individuals in the least deprived quintile. Patients from the most disadvantaged areas were less likely to receive DNase or inhaled antibiotic treatment.

**Interpretation** In the UK, children with cystic fibrosis from more disadvantaged areas have worse growth and lung function compared with children from more affluent areas, but these inequalities do not widen with advancing age. Clinicians consider deprivation status, as well as disease status, when making decisions about treatments, and this might mitigate some effects of social disadvantage.

**Funding** Medical Research Council (UK).

## Introduction

Cystic fibrosis is the most common life-limiting inherited disease in white populations, with most patients dying prematurely from respiratory failure. Children with cystic fibrosis in the UK and in other high-income countries are usually diagnosed in the first year of their life,<sup>1</sup> and subsequently need intensive support from family and health-care services.

Cystic fibrosis is of particular interest in the study of health inequalities, because it is a genetic disease and there is no social gradient in incidence of the disorder—it affects all socioeconomic groups equally (appendix). Inequalities can develop, however, in the outcomes experienced by people with the disease. People with cystic fibrosis from socioeconomically disadvantaged backgrounds, for example, die younger than do those in more advantaged social positions in the UK and the USA.<sup>2-5</sup> Between 1986 to 1994, the adjusted risk of death was 3.65 times higher in patients with cystic fibrosis in the USA with Medicaid cover (taken as an indicator of poverty)

than it was in those without Medicaid cover.<sup>2</sup> In England and Wales, between 1959 and 2008, Barr and Fogarty recorded an increased risk of dying later, at an age above the median of all deaths due to cystic fibrosis, in more advantaged social classes, a pattern that has persisted for more than four decades.<sup>5</sup> As with other chronic diseases, this social patterning of survival in cystic fibrosis implies that social and environmental factors affect outcomes.<sup>6,7</sup> Inequalities in access to specialist health care might also be important, because in many health-care systems provision and use of services decreases with patients' income,<sup>8,9</sup> the so-called inverse care law.<sup>10</sup>

To gain a better understanding of when and how inequalities in outcomes develop in cystic fibrosis, we undertook a longitudinal registry study to explore the effect of deprivation on growth, nutrition, lung function, risk of *Pseudomonas aeruginosa* colonisation, and the use of major cystic fibrosis treatment modalities in a UK-wide population cohort, in the context of a universal health-care system, free at the point of use.

Published Online

January 30, 2013

[http://dx.doi.org/10.1016/S2213-2600\(13\)70002-X](http://dx.doi.org/10.1016/S2213-2600(13)70002-X)

See Online/Comment

[http://dx.doi.org/10.1016/S2213-2600\(13\)70014-6](http://dx.doi.org/10.1016/S2213-2600(13)70014-6)

Department of Public Health and Policy

(D C Taylor-Robinson MBChB, Prof M Whitehead PhD) and Institute of Infection and Global Health

(Prof P J Diggle PhD), University of Liverpool, Liverpool, UK; and UCL Institute of Child Health, London, UK

(Prof R L Smyth FMedSci)

Correspondence to:

Dr David C Taylor-Robinson, Department of Public Health and Policy, Whelan Building, University of Liverpool, Liverpool L69 3GB, UK  
dctr@liv.ac.uk

See Online for appendix

## Methods

### Study design and data sources

We undertook a longitudinal retrospective cohort study of individuals in the UK cystic fibrosis registry who were younger than 40 years at last follow-up, with at least one outcome measurement and a valid postcode between Jan 1, 1996, and Dec 31, 2009. The UK cystic fibrosis registry is supported and coordinated by the UK Cystic Fibrosis Trust.<sup>11,12</sup> The UK cystic fibrosis registry is maintained to a high standard of data quality, and is estimated to include nearly all people thought to have cystic fibrosis in the UK population<sup>13</sup> and is therefore ideally suited to the study of outcomes and treatments across the whole socioeconomic spectrum in the UK (appendix).

NHS research ethics approval (Huntingdon Research Ethics Committee 07/Q0104/2) was granted for the collection of data into the UK database. The Cystic Fibrosis Trust database committee approved the use of anonymised data in this study.

### Primary outcome and covariates

The primary clinical outcomes were weight, height, body-mass index (BMI), percent predicted forced expiratory volume in 1 s (%FEV<sub>1</sub>), and prevalence of *P aeruginosa* colonisation. Anthropometric values were converted into standard deviation [SD] scores using the UK reference population.<sup>14</sup> The primary health-care outcomes were use of treatments in the previous year (yes or no): intravenous antibiotics, supplemental nutritional support, DNase, or inhaled antibiotic treatment. Conditional on the use of intravenous treatment, we also used the log total number of days on intravenous treatment as a secondary outcome.

The primary exposure measure was a small-area-based measure of deprivation of area of residence. Postcodes were used to derive Index of Multiple Deprivation scores for the constituent UK countries<sup>15</sup> and each person was allocated to a deprivation score on the basis of the first recorded postcode on entry to the dataset. Other baseline covariates in the analysis were: sex, genotype coded as the number of delta F508 alleles (0, 1, or 2), year of birth, screening status (diagnosis by neonatal screening or otherwise), and ethnic origin (white or other). Time-varying covariates were age, presence of cystic fibrosis related diabetes (CFRD), and presence of pancreatic insufficiency (ie, whether or not an individual used pancreatic enzyme supplementation). In our health-care use analyses, we adjusted for disease severity on the basis of current %FEV<sub>1</sub>, *P aeruginosa* status, and BMI SD score.

### Statistical analysis

Full details are provided in the supplementary appendix. Briefly, we fitted separate longitudinal models in the paediatric (<18 years) and adult (18–40 years) age ranges. We then approximated time-trends using linear functions (eg, for %FEV<sub>1</sub>), piecewise or broken-stick functions (weight, BMI), or quadratics (eg, any intravenous treatment), as appropriate. For instance, population weight SD score increased to about age 3 years, and then decreased subsequently (appendix). The modelling approach involved first fitting a model adjusted for age and the baseline covariates defined above, and then testing for the significance of adding deprivation. Finally, the time-varying covariates were added to the model, to assess whether the deprivation coefficient was modified. We

	1 (least deprived)	2	3	4	5 (most deprived)	All	p value
Number of patients	1537 (19%)	1563 (19%)	1591 (20%)	1736 (22%)	1628 (20%)	8055	0.0018
Observations (for weight SD score)	9500 (19%)	9706 (20%)	9708 (20%)	10550 (21%)	9873 (20%)	49337	<0.0001
Female sex	712 (46%)	726 (46%)	728 (46%)	825 (48%)	773 (48%)	3764 (47%)	0.38
Age in days at diagnosis (IQR)	121 (30–731)	121 (30–670)	113 (30–730)	109 (30–728)	120 (30–730)	120 (30–730)	0.39
Number of delta F508 alleles							
2	824 (54%)	827 (53%)	822 (52%)	907 (52%)	779 (48%)	4159 (52%)	0.0022
1	543 (35%)	556 (36%)	560 (35%)	609 (35%)	594 (37%)	2862 (36%)	0.63
0	170 (11%)	180 (12%)	209 (13%)	220 (13%)	255 (16%)	1034 (13%)	<0.0001
Non-white	31 (2%)	31 (2%)	52 (3%)	73 (4%)	120 (7%)	307 (4%)	<0.0001
Screened	233 (15%)	272 (17%)	245 (15%)	282 (16%)	277 (17%)	1309 (16%)	0.39
Birth cohort							
1957 to 1966	62 (4%)	49 (3%)	64 (4%)	51 (3%)	35 (2%)	261 (3%)	<0.0045
1967 to 1976	157 (10%)	172 (11%)	182 (11%)	171 (10%)	153 (9%)	835 (10%)	0.23
1977 to 1986	329 (21%)	384 (25%)	369 (23%)	426 (25%)	396 (24%)	1904 (24%)	0.09
1987 to 1996	496 (32%)	478 (31%)	489 (31%)	535 (31%)	530 (33%)	2528 (31%)	0.82
1997 to 2006	396 (26%)	393 (25%)	396 (25%)	427 (25%)	410 (25%)	2022 (25%)	0.62
2007 to <2010	97 (6%)	87 (6%)	91 (6%)	126 (7%)	104 (6%)	505 (6%)	0.32

Data are n (%) unless otherwise stated.

**Table 1: Unadjusted characteristics of study population by deprivation quintile (UK cystic fibrosis registry 1996 to 2009)**

estimated all model parameters by maximum likelihood, using linear or generalised linear mixed effects models.<sup>16</sup> We used generalised likelihood ratio statistics to compare nested models, and Wald statistics to test hypotheses about model parameters. We used R (version 2.9.2) for all statistical analyses.

### Role of the funding source

The study sponsor had no role in the design, collection, analysis, or interpretation of the data, in the writing of the report, or in the decision to submit the article for publication. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

### Results

The final dataset for weight SD scores, the most commonly collected outcome, contained information collected at 49 337 annual reviews for 8055 patients between Jan 1, 1996 and Dec 31, 2009, in the UK (table 1 and appendix). 5324 (66%) individuals had five or more follow-up measures (appendix), with a total of 48 425 person-years of follow-up. We recorded no relation between sex ratios, birth cohort, neonatal screening and deprivation status (table 1), or number of incident cases, and age at diagnosis and deprivation status (appendix). We recorded a slight trend towards fewer heterozygote delta F508 carriers ( $p=0.0022$ ), more people with no delta F508 genes ( $p<0.0001$ ), and a greater proportion of non-white patients with increasing level of deprivation ( $p<0.0001$ ). Compared with the UK reference population, the population of patients with cystic fibrosis weighed less (SD score  $-0.37$ , 95% CI  $-0.43$  to  $-0.35$  [35th centile]), were shorter ( $-0.50$ ,  $-0.53$  to  $-0.47$  [30th centile]), and had a lower BMI ( $-0.08$ ,  $-0.11$  to  $-0.06$  [46th centile]; in models ignoring time trends).

Weight SD scores increased from diagnosis up to about the age of 3 years, decreasing thereafter (appendix). After adjustment for baseline factors, at diagnosis, the weight of children in the most deprived quintile was lower than that of children in the least deprived quintile (weight SD score  $-0.54$ , 95% CI  $-0.73$  to  $-0.34$ ). The deprivation gap diminished with increasing age up to age 3 years, and from then on remained constant (table 2 and appendix). A higher weight SD score was associated with male sex, screened patients, heterozygotes for delta F508, and white patients (appendix). In adults, adjusted weight-for-age was lower in more deprived groups (table 2).

The average height of individuals in the most deprived quintile compared with the least deprived quintile was also about a third of an SD score shorter in the adjusted analysis, a difference that remained constant across all ages (table 2 and appendix). In patients younger than 18 years, a bigger height SD score was statistically significantly associated with male sex and screened patients, and statistically significantly increased in white patients with age (appendix).

	Patients younger than 18 years	Patients aged 18 years to <40 years
<b>Clinical outcomes*</b>		
FEV <sub>1</sub> (percentage points [95% CI])	-4.12 (-5.01 to -3.19)	-1.6 (-4.41 to 1.25)
Weight-for-age (SD score [95% CI])	-0.28 (-0.38 to -0.18)	-0.31 (-0.46 to -0.16)
Height-for-age (SD score [95% CI])	-0.31 (-0.40 to -0.21)	-0.31 (-0.43 to -0.19)
BMI-for-age (SD score [95% CI])	-0.13 (-0.22 to -0.04)	-0.12 (-0.25 to 0.01)
<i>Pseudomonas aeruginosa</i> colonisation (OR [95% CI])	1.89 (1.34 to 2.66)	1.78 (1.26 to 2.51)
<b>Treatments</b>		
Any intravenous treatment (OR [95% CI])†	2.52 (1.92 to 3.17)	1.89 (1.51 to 2.38)
Total intravenous days per year (% change [95% CI])†	15.9 (8.2 to 24)	10.6 (2.5 to 19.2)
Supplemental feeding (OR [95% CI])‡	1.78 (1.42 to 2.2)	2.38 (1.69 to 3.36)
DNase treatment (OR [95% CI])†	0.40 (0.21 to 0.72)	0.37 (0.26 to 0.52)
Use of inhaled antibiotics (OR [95% CI])†	0.66 (0.47 to 0.93)	0.40 (0.31 to 0.5)

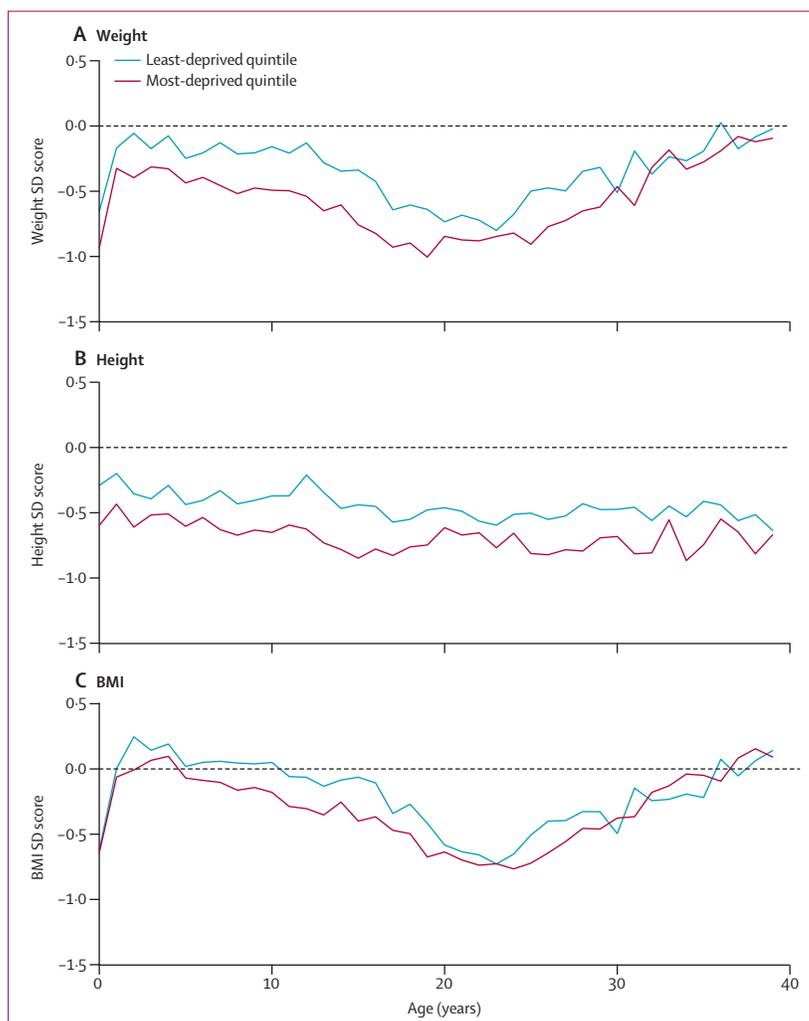
All estimates compare the most deprived quintile to the least deprived (reference) quintile. \*The outcomes are from separate longitudinal models adjusted for time trends, sex, genotype, screening status, and ethnic origin. †Adjusted for time trends, sex, genotype, screening status, (FEV<sub>1</sub>), and *Pseudomonas aeruginosa* colonisation status. ‡Adjusted for time trends, sex, genotype, screening status, and body mass index (BMI) SD score.

**Table 2: Summary of adjusted effects of deprivation on clinical outcomes and use of treatments in patients with cystic fibrosis in the UK**

We modelled BMI SD score much like we modelled weight SD score, with a split-line at age 3 years. In the paediatric age range, there was a deprivation gap (with lower scores in the most deprived groups) of  $-0.13$  ( $-0.22$  to  $-0.04$ ; table 2). Higher BMI was associated with male sex in the paediatric age range (individuals ages 0–18 years), and had a steeper rate of decline in delta F508 homozygotes after the age of 3 years (appendix). In the adult age range, we recorded no association between BMI SD score and deprivation status ( $-0.12$ ,  $-0.25$  to  $0.01$ ; figure 1 and table 2).

Addition of the time-varying covariates did not substantially alter the deprivation effects for growth outcomes, and the estimates were consistent with a monotonic dose-response relation between deprivation and both weight and height.

In the final model for %FEV<sub>1</sub>, we detected a difference of  $-4.1$  percentage points ( $-5.0$  to  $-3.1$ ) when comparing children (<18 years) in the most deprived quintile with those in the least deprived quintile (a difference that was apparent from as soon as %FEV<sub>1</sub> can be measured at about 5 years of age), but there was no evidence of an increased rate of decline in children from more deprived quintiles. Higher %FEV<sub>1</sub> was associated with male sex, screened patients, heterozygote delta F508 status, white patients, no CFRD, no *P aeruginosa* colonisation, and higher BMI (figure 2 and appendix). Further adjustment for *Burkholderia cenocepacia* status and care centre did not change the deprivation effect on %FEV<sub>1</sub> (appendix). The addition of BMI SD score to the model reduced the %FEV<sub>1</sub> deprivation gap to  $-3.5$  percentage points ( $-5.2$  to  $-1.8$ ). There was no statistically significant association between %FEV<sub>1</sub> and social deprivation in the adult age range (table 2). The cross-sectional proportion of people with chronic *P aeruginosa* infection increased steadily with age



**Figure 1:** Comparison of anthropometric outcomes, by age and socioeconomic status  
Mean cross-sectional (A) weight, (B) height, and (C) body-mass index (BMI).

to about 60% by the age of 20 years, and was more common in the most deprived quintile, with an odds ratio (OR) of 1.9 (95% CI 1.3 to 2.7) in the adjusted paediatric analysis for the most deprived quintile (table 2 and figure 2). An increased likelihood of *P aeruginosa* colonisation was associated with female sex, homozygote delta F508 status, CFRD, pancreatic insufficiency, and lower %FEV<sub>1</sub>, but adjustment for these factors did not substantially alter the deprivation effect (data not shown). The estimates were consistent with a monotonic dose-response relation between deprivation and %FEV<sub>1</sub> (appendix) and risk of *P aeruginosa* colonisation (data not shown).

The use of any intravenous treatment, after adjustment for disease severity, was more than twice as common in the most deprived children cohort compared with the least deprived children cohort (table 2), and this deprivation difference was also present in adults (table 2 and figure 3). Further adjustment for care centre did not change this effect. Conditional on receipt of intravenous treatment,

and after adjustment for disease severity, people in the most deprived quintile had more days of intravenous treatment in both the paediatric and adult age range (table 2). We analysed the receipt of hospital and at-home intravenous treatment separately and noted that the higher prevalence of any intravenous treatment seen in the most deprived quintile was almost entirely due to delivery of such treatment in hospital rather than home (figure 3). For intravenous treatment at home, the association with social deprivation was much less strong, and, in the cross-sectional analysis, at-home intravenous treatment was more common in the least deprived quintile compared with the most deprived quintile in patients between the ages of 10 years and 27 years (figure 3). Prevalence of any supplemental feeding therapy in the previous year was more common in the most deprived quintile, compared to the least, across the entire age range from age 0 years to age 40 years (OR 1.78, 95% CI 1.42 to 2.2, adjusted for baseline variables, *P aeruginosa* infection status, and BMI, in the 5–18 age group, figure 3).

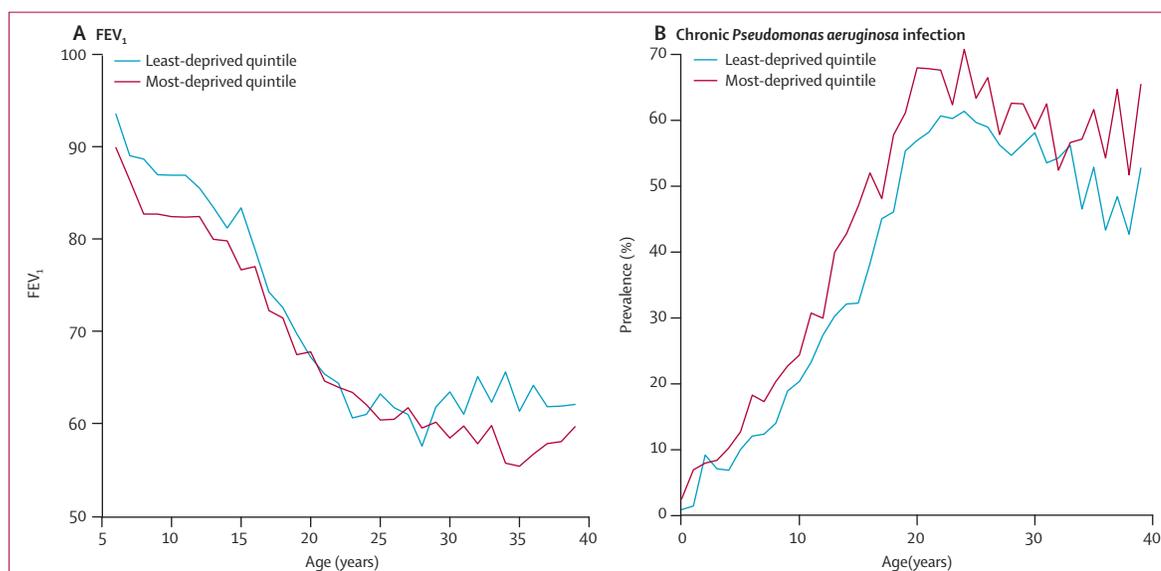
We detected no statistically significant association between DNase use and deprivation in the paediatric age range before we adjusted for disease severity. After adjustment for disease severity, treatment was less likely in the most deprived quintile, in both children and adults, although the association with deprivation was stronger in adults. We saw a similar pattern for inhaled antibiotic treatment (table 2 and figure 3). The estimates were consistent with a monotonic dose-response relation between socioeconomic status and treatment outcomes (appendix).

## Discussion

Our findings show that children with cystic fibrosis from the most disadvantaged areas in the UK have lower weight, height, and BMI in the first years of life after diagnosis, are more likely to have chronic *P aeruginosa* infection, and have a lower %FEV<sub>1</sub> than do children in the least disadvantaged areas. These social inequalities persist into adulthood but do not widen.

Our findings suggest evidence of positive discrimination, or so-called pro-poor bias, in the provision of some key treatments, on the basis of socioeconomic circumstances. We show that in the NHS, compared with children with cystic fibrosis in the least disadvantaged areas, children with cystic fibrosis from the most disadvantaged areas are about twice as likely, after adjustment for disease severity, to receive intravenous antibiotics (specifically in hospital) and nutritional support. Our findings also show some apparent bias in favour of wealthier populations, a so-called pro-rich bias, in two other treatments, DNase and inhaled antibiotics, with patients from the most affluent areas being more likely to receive these treatments after adjustment for disease severity.

Key strengths of this study include the population-wide coverage of the UK cystic fibrosis registry, the high quality of the data, and the longitudinal analysis. However, our study does have limitations. First, it relies on



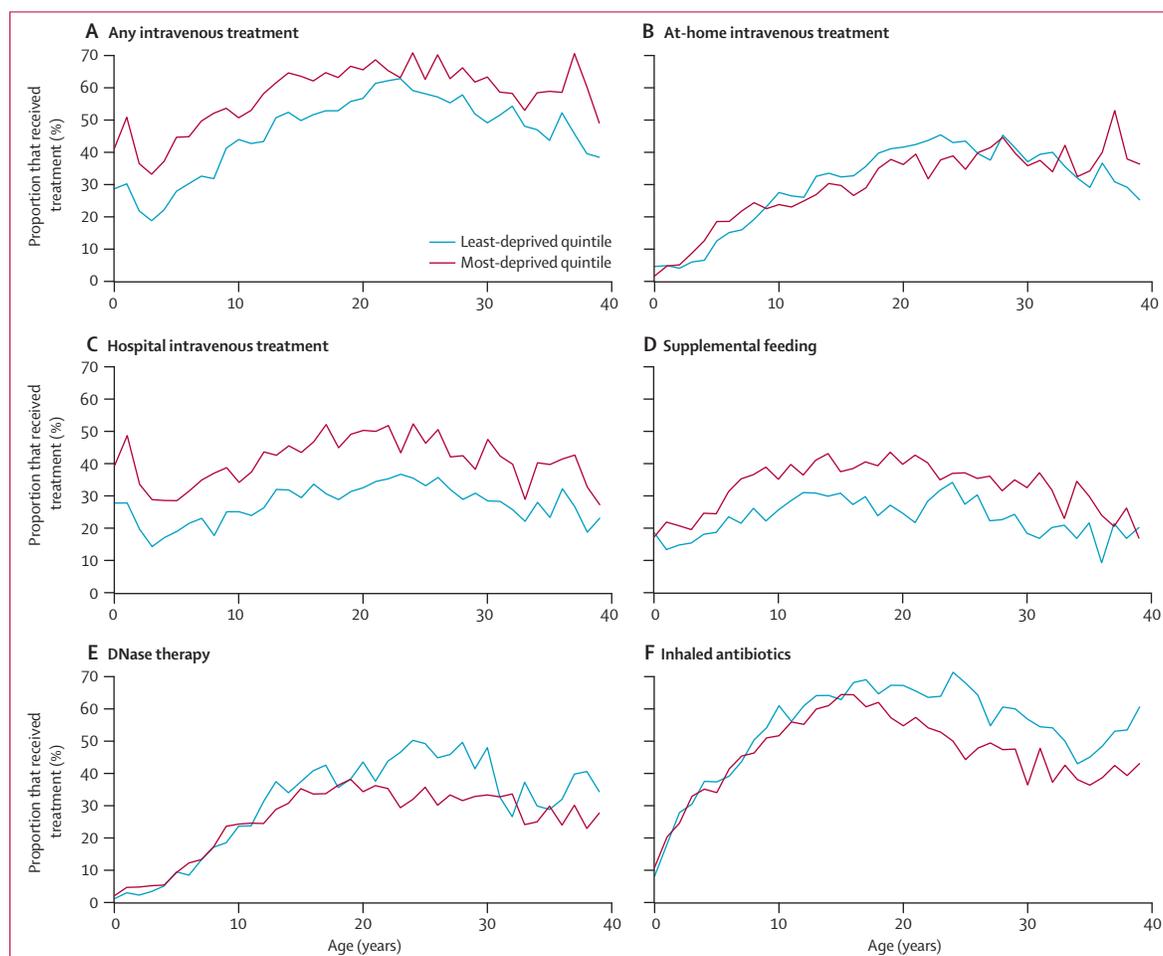
**Figure 2: Comparison of respiratory outcomes, by age and socioeconomic status**  
Mean cross-sectional (A) FEV<sub>1</sub> and (B) *Pseudomonas aeruginosa* colonisation prevalence.

retrospective, routinely collected data and we used a standard measure of deprivation of area of residence. Each small area contains about 1500 people, and, in this respect, the Index of Multiple Deprivation scores allowed much finer resolution than US analyses<sup>3,17,18</sup> that have used ZIP-code-linked income data, because every ZIP code contains about 30 000 people (panel).<sup>19</sup> There is always the possibility of ecological fallacy (whereby inferences made at the group level do not apply to the individual), but this possibility is unlikely in view of the fact that similar associations have been seen in the US studies that use both area and individual measures of socioeconomic status.<sup>3,17,18</sup> Second, we had valid postcodes for only 90% of the sample, although our sample size was large, with no pronounced gradient in the proportion of patients by deprivation quintile. The excluded population—those with no valid postcode—were largely older birth cohorts, owing to the improved collection of postcodes by clinical staff over time, but we do not believe that this has biased the associations detected in our analysis (appendix). Third, there is a strong cohort effect in cystic fibrosis and, with datasets of this type, age and cohort effects confound one another, and cannot be completely separated.<sup>20</sup> We have adjusted for both in our analysis, to estimate the adjusted effect on deprivation.

Overall, the UK cystic fibrosis population is underweight and shorter compared with the UK reference population, by about a third of an SD score. Deprivation roughly doubles this effect, lowering the SD score by another third. How much of the effect of socioeconomic status on growth outcomes is specific to cystic fibrosis, and how much is attributable to socioeconomic status effects in the general population is unclear. Comparable data in contemporary representative cohorts in the UK is

absent, but the age-related changes in growth in the general population are characterised by increasing obesity in childhood from the age of 4 years onwards, with higher BMI in the more deprived populations,<sup>21,22</sup> findings which contrast with the patterns seen in our study. The projected weight difference at intercept in our study ( $-0.54$ ) by socioeconomic status is also larger than those in other recent studies,<sup>23,24</sup> but direct comparison between these studies and ours is complicated by the use of different socioeconomic status measures. We speculate that having cystic fibrosis is likely to amplify the effects of socioeconomic status on nutritional status at birth and in the first few years of an individual's life.

The inequality in weight is greatest at around the time of diagnosis, and becomes narrower over the first 3 years of life. This is an important finding, because a widening of inequalities over time is often the norm.<sup>7,25,26</sup> These findings suggest that extending the period of differential weight gain for as long as possible might reduce inequalities, further lending support to neonatal screening programmes to enable early diagnosis and treatment.<sup>27</sup> We speculate that by extending this period of catch-up for as long as possible by early diagnosis (ie, screening) we might see an attenuation of the deprivation effect over time. In this study, we detected no difference in the age at diagnosis by deprivation, but screening was associated with increased weight and height and improved lung function in children. Furthermore, our finding that the prevalence of supplemental feeding treatment was higher, after adjusting for disease severity, in the most disadvantaged patients suggests that NHS professionals are actively engaged in trying to boost the nutrition of poorer patients, recognising their health disadvantage.



**Figure 3: Comparison of treatment methods, by age and socioeconomic status**

Proportion of patients who received (A) any intravenous antibiotic treatment, (B) home intravenous antibiotic treatment, (C) hospital intravenous antibiotic treatment, (D) supplemental feeding, (E) DNase, and (F) inhaled antibiotics.

The socioeconomic gradient in lung function, evident as soon as it can be routinely measured at the age of 5 years, points to the crucial role of environmental and health-care factors operating in the early years of life to produce inequalities. It further reinforces the need for early diagnosis and action to prevent adverse consequences for children with cystic fibrosis living in disadvantaged circumstances. In Schechter and colleagues' cross-sectional study of US data,<sup>2</sup> inequalities in %FEV<sub>1</sub> by Medicaid status widened slightly from age 5 years to age 20 years. The magnitude of the inequalities in lung function at the age of 5 years seen in Schechter's study was larger (about a 9% difference) than in our UK study (4%), as was the magnitude of inequalities in lung function seen in O'Connor and colleagues' US study,<sup>3</sup> which showed a difference of 5.5% between the most and least deprived quintiles. Methodological differences between the studies, however, make a direct comparison between these UK and US findings inappropriate. This study is the first to examine the relation between

deprivation and %FEV<sub>1</sub> in a population-level, adult cohort. We did not detect an association, despite the higher prevalence of *P aeruginosa*. We speculate that this finding might relate to the complication of progressive drop-out in older patients, and the insensitivity of %FEV<sub>1</sub> as an outcome measure in adults.<sup>20</sup>

The increased prevalence of chronic *P aeruginosa* infection in patients from more deprived areas, after adjusting for %FEV<sub>1</sub>, is a new finding in a population-level cohort. In Schechter and colleagues' study,<sup>2</sup> Medicaid-insured patients were more likely to have *P aeruginosa* infection than were patients who were not eligible for Medicaid insurance, but when adjusted for %FEV<sub>1</sub> there was no statistically significant difference—another US cohort study did not detect an association either.<sup>28</sup> Previously identified risk factors for *P aeruginosa* acquisition, which is associated with worse lung function, include female sex and genotype (both associations shown in this study), and exposure to other patients with *P aeruginosa* colonisation.<sup>29</sup> Our finding that more deprived

groups are more likely to receive intravenous treatment in hospital might result in more deprived patients having greater exposure to other patients with chronic *P aeruginosa*, therefore increasing their risk of infection.

We saw substantial socioeconomic differences in the reported use of key cystic fibrosis treatments in two contrasting ways. First, children from the most deprived quintile were about twice as likely to receive hospital intravenous antibiotic treatment and nutritional support, after adjustment for disease severity, compared with those from the least deprived quintile. We can speculate, from our knowledge of UK cystic fibrosis services, that clinicians in the NHS are more likely to bring children from more deprived areas into hospital for intravenous treatment because of concerns about the difficulties in delivering treatments in their homes. Conversely, children living in more affluent areas might receive intravenous treatment at home because of judgments about the adequacy of support and adherence to treatment in their home or because of their families' wish to avoid disruption to schooling and family life. This equitable model of care, with positive discrimination for socially disadvantaged children and adults with cystic fibrosis, is an uncommon finding in health systems, when access, particularly to secondary care for adults, often decreases with increasing deprivation, after adjusting for differential need.<sup>8,30</sup> While several studies have seen use of health services by level of deprivation to be more equal in relation to children than adults,<sup>31</sup> we have detected evidence in children with cystic fibrosis that goes even further with a pro-poor bias in the NHS for specific treatments. Coupled with our findings of inequalities in outcomes by deprivation, which do not widen over time, we speculate that the treatment decisions being made by clinicians might mitigate some effects of social disadvantage. This provides encouragement that there are interventions that health services can make to reduce the adverse effects of deprivation on chronic disorders such as cystic fibrosis. In the USA, with use of ZIP-code-linked income of an area as the socioeconomic indicator, there was no gradient in intravenous treatment use in children younger than 12 years, but in young people aged 13–18 years, those living in more affluent areas were more likely to be treated (13·8% in the lowest income category compared with 19·2% in the highest).<sup>18</sup>

Our second, and contrasting, set of findings on cystic fibrosis treatments, however, point to an apparent pro-rich bias in two other treatments, which were more evident in adults than in children: more affluent adults in the UK were more likely to receive DNase and inhaled antibiotics than were their more disadvantaged counterparts. DNase is an expensive treatment to reduce viscosity of sputum and to aid sputum expectoration, and some evidence exists that it prevents decrease in %FEV<sub>1</sub>.<sup>32</sup> These treatments, although expensive, are free of charge to all patients in the NHS. One possibility for the social disparity in access to them is that they are both home-based treatments, requiring regular and long-term

administration. Socially disadvantaged patients with cystic fibrosis are less likely to adhere to treatments,<sup>29</sup> and if they report poor adherence, clinicians might be less likely to prescribe these drugs because they are unlikely to be as effective if taken inconsistently. Evidence from the USA shows no difference in use of DNase in children by area income quintile, but Medicaid-insured children (ie, those receiving free or subsidised care) were more likely to receive DNase than were children who were not eligible for Medicaid insurance.<sup>17</sup>

Further research is needed to clarify which elements of the cystic fibrosis care model might contribute to a reduction in the adverse outcomes associated with deprivation. A cause for concern is the fact that the most disadvantaged families have a higher burden of treatment, in terms of time spent in hospital, which increases disruption to school and family life. Furthermore, the link with *P aeruginosa* colonisation requires further investigation. Higher socioeconomic status, as measured by parental education status, is associated with improved adherence to treatment in cystic fibrosis,<sup>29</sup> and further research is needed to investigate the processes that lead to these differences. Systems to support the provision of intravenous treatment at home for more deprived groups in the UK should be explored.

Differences in access to health care cannot explain the differences in weight and height, by socioeconomic status, that are evident at the time of diagnosis, and are

#### Panel: Research in context

##### Systematic review

We searched PubMed with the terms "(cystic fibrosis) and (inequality OR equity OR inequity OR socioeconomic OR disadvantage OR vulnerable OR poverty OR social class OR disparity)" to identify relevant studies on the effect of socioeconomic status on outcomes and treatment in people with cystic fibrosis. We applied no date or language restrictions. We identified a review that summarises all studies,<sup>29</sup> much of which were done in the USA, where the health-care system is different to that in the UK. People with cystic fibrosis from socioeconomically disadvantaged backgrounds die younger than do those in more advantaged social positions in the UK<sup>5</sup> and the USA.<sup>2</sup> The key challenge is to understand how and when these inequalities develop, and to understand how the health-care system in the UK can mitigate or perpetuate these effects to identify promising options for intervention.

##### Interpretation

This study has identified important longitudinal differences in weight, height, body-mass index, forced expiratory volume in 1 s, and risk of *Pseudomonas aeruginosa* colonisation by deprivation in people with cystic fibrosis in the UK, which start early in life, but do not increase over time. We detected socioeconomic differences in the reported use of key treatments in the UK. People from more deprived areas are about twice as likely to receive in-hospital intravenous antibiotic treatment and nutritional support, but less likely to receive DNase and inhaled antibiotics. Interventions to reduce inequalities in outcomes in cystic fibrosis need to be focused in the antenatal period and the early years of life. Such interventions include smoking prevention and public health initiatives to address inequalities in maternal and child health. Further research is needed to clarify which elements of the cystic fibrosis care model in the UK might contribute to a reduction in the adverse outcomes associated with deprivation, and to investigate identified differences in access to inhaled treatments.

unlikely to explain the gradient in lung function evident at around the age of 5 years. The UK cystic fibrosis registry does not capture data about smoking in the home and these early effects might be associated with the known differences in smoking prevalence by socioeconomic status in the UK.<sup>33</sup> The effect of socioeconomic status on growth in utero and in the early years of life in people with cystic fibrosis, might be mediated, at least in part, by maternal smoking, thus affecting subsequent outcomes and ultimately survival.

Future studies should focus on the assessment of interventions, such as the reduction of exposure to environmental tobacco smoke,<sup>34</sup> which might mitigate the effects of deprivation during the critical early years of life, and on the identification of aspects of health-care provision in cystic fibrosis that would help overcome the extra burden of adverse consequences of cystic fibrosis faced by patients living in economically-disadvantaged circumstances.

#### Contributors

DCT-R, PJD, MW, and RLS had the idea for and designed the study and were named on the original MRC Fellowship application. DCT-R undertook the analysis and PJD supervised analysis. DCT-R, MW, RLS, and PJD interpreted the results and drafted the paper. All authors contributed to and approved the final draft for publication.

#### Conflicts of interest

We declare that we have no conflicts of interest.

#### Acknowledgments

DCT-R is supported by an MRC Population Health Scientist Fellowship (G0802448). We thank the UK Cystic Fibrosis Trust for access to the UK cystic fibrosis registry and all the centre directors for the input of data to the registry. Elaine Gunn assisted with access to the data.

#### References

- Cystic Fibrosis Trust. UK CF Registry annual data report 2010. <http://www.cftrust.org.uk/aboutcf/publications/cfregistryreports/> (accessed Sept 28, 2012).
- Schechter MS, Shelton BJ, Margolis PA, Fitzsimmons SC. The association of socioeconomic status with outcomes in cystic fibrosis patients in the United States. *Am J Respir Crit Care Med* 2001; **163**: 1331–37.
- O'Connor GT, Quinton HB, Kneeland T, et al. Median household income and mortality rate in cystic fibrosis. *Pediatrics* 2003; **111**: e333–39.
- Britton JR. Effects of social class, sex, and region of residence on age at death from cystic fibrosis. *BMJ* 1989; **298**: 483–87.
- Barr HL SA, Fogarty AW. The association between socioeconomic status and gender with median age at death from cystic fibrosis in England and Wales: 1959 to 2008. *BMJ* 2011; **343**: d4662.
- Marmot MJ, Goldblatt P, Boyce T, McNeish D, Grady M, Geddes I. Fair society, healthy lives: strategic review of health inequalities in England post-2010. London: The Marmot Review, 2010.
- Ben-Shlomo Y, Kuh D. A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *Int J Epidemiol* 2002; **31**: 285–93.
- Stirbu I, Kunst AE, Mielck A, Mackenbach JP. Inequalities in utilisation of general practitioner and specialist services in 9 European countries. *BMC Health Serv Res* 2011; **11**: 288.
- van Doorslaer E, Masseria C, Koolman X. Inequalities in access to medical care by income in developed countries. *CMAJ* 2006; **174**: 177–83.
- Hart JT. The inverse care law. *Lancet* 1971; **1**: 405–12.
- Cystic Fibrosis Trust. UK CF Registry. <http://www.cftrust.org.uk/aboutcf/publications/cfregistryreports/> (accessed Jan 15, 2013).
- Adler AI, Shine BS, Chamnan P, Haworth CS, Bilton D. Genetic determinants and epidemiology of cystic fibrosis-related diabetes: results from a British cohort of children and adults. *Diabetes Care* 2008; **31**: 1789–94.
- Mehta G, Sims EJ, Culross F, McCormick JD, Mehta A. Potential benefits of the UK Cystic Fibrosis Database. *J R Soc Med* 2004; **97**: 60–71.
- Pan H, Cole T. User's guide to LmsGrowth. London: Medical Research Council, 2005.
- Geoconvert. Welcome to GeoConvert. <http://geoconvert.mimas.ac.uk/> (accessed Dec 22, 2012).
- Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982; **38**: 963–74.
- Schechter MS, McColley SA, Silva S, Haselkorn T, Konstan MW, Wagener JS. Association of socioeconomic status with the use of chronic therapies and healthcare utilization in children with cystic fibrosis. *J Pediatr* 2009; **155**: 634–39.
- Schechter MS, McColley SA, Regelman W, et al. Socioeconomic status and the likelihood of antibiotic treatment for signs and symptoms of pulmonary exacerbation in children with cystic fibrosis. *J Pediatr* 2011; **159**: 819–24.
- Krieger N, Waterman P, Chen JT, Soobader MJ, Subramanian SV, Carson R. Zip code caveat: bias due to spatiotemporal mismatches between zip codes and US census-defined geographic areas—the Public Health Disparities Geocoding Project. *Am J Public Health* 2002; **92**: 1100–02.
- Taylor-Robinson D, Whitehead M, Diderichsen F, et al. Understanding the natural progression in %FEV1 decline in patients with cystic fibrosis: a longitudinal study. *Thorax* 2012; **67**: 860–66.
- Howe LD, Tilling K, Galobardes B, Smith GD, Gunnell D, Lawlor DA. Socioeconomic differences in childhood growth trajectories: at what age do height inequalities emerge? *J Epidemiol Community Health* 2012; **66**: 143–48.
- Howe LD, Tilling K, Galobardes B, Smith GD, Ness AR, Lawlor DA. Socioeconomic disparities in trajectories of adiposity across childhood. *Int J Pediatr Obes* 2011; **6**: e144–53.
- Kinra S, Baumer JH, Davey Smith G. Early growth and childhood obesity: a historical cohort study. *Arch Dis Child* 2005; **90**: 1122–27.
- Wijlaars LP, Johnson L, van Jaarsveld CH, Wardle J. Socioeconomic status and weight gain in early infancy. *Int J Obes (Lond)* 2011; **35**: 963–70.
- CSDH. Closing the gap in a generation: health equity through action on the social determinants of health. Final report of the commission on social determinants of health. [http://whqlibdoc.who.int/publications/2008/9789241563703\\_eng.pdf](http://whqlibdoc.who.int/publications/2008/9789241563703_eng.pdf) (accessed Jan 15, 2013).
- Whitehead M, Dahlgren G. Concepts and principles for tackling social inequities in health: levelling up part 1. Copenhagen: WHO, 2007.
- Southern KW, Merelle MM, Dankert-Roelofs JE, Nagelkerke AD. Newborn screening for cystic fibrosis. *Cochrane Database Syst Rev* 2009; **1**: CD001402.
- Rosenfeld M, Emerson J, McNamara S, et al. Risk factors for age at initial *Pseudomonas* acquisition in the cystic fibrosis epic observational cohort. *J Cyst Fibros* 2012; **11**: 446–53.
- Schechter MS. Nongenetic influences on cystic fibrosis outcomes. *Curr Opin Pulm Med* 2011; **17**: 448–54.
- Hanratty B, Zhang T, Whitehead M. How close have universal health systems come to achieving equity in use of curative services? A systematic review. *Int J Health Serv* 2007; **37**: 89–109.
- Groholt EK, Stigum H, Nordhagen R, Kohler L. Health service utilization in the Nordic countries in 1996: influence of socio-economic factors among children with and without chronic health conditions. *Eur J Public Health* 2003; **13**: 30–37.
- Pressler T. Review of recombinant human deoxyribonuclease (rhDNase) in the management of patients with cystic fibrosis. *Biologics* 2008; **2**: 611–17.
- Mackenbach JP. What would happen to health inequalities if smoking were eliminated? *BMJ* 2011; **342**: d3460.
- Taylor-Robinson D, Schechter MS. Health inequalities and cystic fibrosis. *BMJ* 2011; **343**: d4818.

# A Longitudinal Study of the Impact of Social Deprivation and Disease Severity on Employment Status in the UK Cystic Fibrosis Population

David C. Taylor-Robinson<sup>1\*</sup>, Rosalind Smyth<sup>2</sup>, Peter J. Diggle<sup>3</sup>, Margaret Whitehead<sup>1</sup>

<sup>1</sup> Department of Public Health and Policy, University of Liverpool, Liverpool, United Kingdom, <sup>2</sup> Institute of Child Health, UCL, London, United Kingdom, <sup>3</sup> Institute of Infection and Global Health, University of Liverpool, Liverpool, United Kingdom

## Abstract

**Background:** People with Cystic Fibrosis (CF) in the UK and elsewhere are increasingly surviving into adulthood, yet there is little research on the employment consequences of having CF. We investigated, for the first time in a UK-wide cohort, longitudinal employment status, and its association with deprivation, disease severity, and time in hospital.

**Methods:** We did a longitudinal registry study of adults with CF in the UK aged 20 to 40 (3458 people with 15,572 observations between 1996 and 2010), using mixed effects models.

**Results:** Around 50% of adults with CF were in employment. Male sex, higher lung function and body mass index, and less time in hospital were associated with improved employment chances. All other things being equal, being in the most deprived quintile was associated with a reduction of employment prevalence of 17.6 percentage points compared to the prevalence in the least deprived quintile. Having poor lung function was associated with a reduced employment prevalence of 7.2 percentage points compared to the prevalence for people with relatively good lung function. Acting synergistically, deprivation modifies the effect of lung function on employment chances – poor lung function in the least deprived group was associated with a 3 percentage point reduction in employment chances, while poor lung function in the most deprived quintile was associated with a 7.7 point reduction in employment chances.

**Conclusions:** Greater deprivation, disease severity, and time in hospital are all associated with employment chances in adults with CF. Furthermore, our analysis suggests that deprivation amplifies the harmful association of disease severity on employment. Future studies should focus on understanding and mitigating the barriers to employment faced by people with CF.

**Citation:** Taylor-Robinson DC, Smyth R, Diggle PJ, Whitehead M (2013) A Longitudinal Study of the Impact of Social Deprivation and Disease Severity on Employment Status in the UK Cystic Fibrosis Population. PLoS ONE 8(8): e73322. doi:10.1371/journal.pone.0073322

**Editor:** Shama Ahmad, University of Colorado, Denver, United States of America

**Received:** May 7, 2013; **Accepted:** July 18, 2013; **Published:** August 23, 2013

**Copyright:** © 2013 Taylor-Robinson et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** DTR is supported by an MRC Population Health Scientist Fellowship (G0802448). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: dctr@liv.ac.uk

## Introduction

Cystic fibrosis (CF) is the commonest life-limiting inherited disease among Caucasian populations, with most patients dying prematurely from respiratory failure. Children with CF in the UK and other high-income countries are usually diagnosed in the first year of life [1], and subsequently require intensive support from family and healthcare services. People with CF in the UK and elsewhere are increasingly surviving into adulthood, with the median age of survival estimated to be over fifty years for a person born in this century [2]. One implication of this improved survival, is that increasing attention needs to be paid to the experiences of people with CF when they reach adulthood and enter employment.

Employment is one of the “social determinants” of health [3]. Work influences health in a number of ways; it provides income to meet material needs, but also fulfils critical psycho-social functions, increasing self-worth, sense of identity and opportunities for social

interaction. Numerous studies have identified unemployment as a potent risk factor for poor health, and equally, poor health can lead to reduced employment chances [4,5]. People with chronic illnesses face numerous barriers to entering the labour market, and CF provides a case in point. Factors related to disease severity, such as reduced lung function may restrict employment choices for adults with CF, and the treatment burden further compounds this: adults with CF are generally expected to perform physiotherapy regularly and there are the added demands of taking large numbers of therapies, including frequent visits to hospital [6].

Despite these potential challenges, the evidence about patterns of employment for adults with CF is limited [7], and mainly based on cross-sectional studies of single centres, which cannot delineate the relationships between chronic illness and employment outcomes or whether these relationships indicate causality. Furthermore, Edwards et al [8], adopting the social model of disability, have criticised the approach taken to understanding employment outcomes in CF. They point out that most of the

research, to date, has been restricted to the effects of disease severity on employment chances, ignoring the significant structural and societal barriers to employment for people with chronic illness.

Exploring inequalities by socio-economic status (SES) in employment outcomes in people with CF is a key step in understanding how health and social inequalities are generated and perpetuated. Because CF is a classically inherited genetic disease, unlike most chronic diseases, there is no difference in the incidence of the condition with socioeconomic status [9,10]. However, inequalities develop over the course of people's lives, as a consequence of having the disease. Informed by Diderichsen's analytic framework of the pathways from social context to health outcomes [11], we have demonstrated that in the UK CF population there are clinically important differences in growth, and lung function by deprivation, which are evident early on in children's lives [9,10]. Furthermore, people with CF from socio-economically disadvantaged backgrounds die at a younger age than those in more advantaged social positions in the UK and the US [12–15]. The social patterning of outcomes in cystic fibrosis implies that the mechanism of *differential exposure* to social and environmental risk factors is playing an important role in influencing outcomes [3,11,16].

Building on these findings, the next step is to look for any “differential social consequences” of ill-health in the context of CF [11]. Our aims in this study were to explore the association of prior deprivation, disease severity, and time in hospital on longitudinal employment chances in people with CF and to investigate whether changes in lung function have differential effects on employment chances by deprivation (*‘differential social consequences’ in Diderichsen’s model*). For instance, is poor lung function in CF more damaging to employment chances in people from more disadvantaged areas? We undertook a longitudinal population level registry study of employment status in adults with CF in the UK to address this question.

## Methods

### Ethics statement

NHS research ethics approval (Huntingdon Research Ethics Committee 07/Q0104/2) has been granted for the collection of data into the UK database. Each patient provided written informed consent for collection of data in the registry, and for use of anonymised data in research. The CF Trust database committee approved the use of anonymised data in this study, under the terms of the NHS ethics approval.

### Design, setting and data source

We undertook a longitudinal retrospective cohort study of annual review data on individuals between the ages of 20 and 40 with at least one outcome measurement and a valid postal code in the UK CF Registry between 1996 and 2010. The UK CF Registry, co-ordinated by the Cystic Fibrosis Trust [17,18], is maintained to a high standard of data quality, and includes nearly all people with CF in the UK population [19], with an estimated coverage of over 99% [9], and is therefore ideally suited to the study of prior exposures on subsequent employment outcomes across the whole socioeconomic spectrum in the UK society.

### Primary outcome and covariates

The primary longitudinal outcome was any employment (defined as “full” or “part-time” as specified in the drop-down menu in the UK CF registry) in the preceding year (yes or no) recorded at annual review. The primary exposure measures were

deprivation of small-area of residence, as commonly used in epidemiological studies in the UK [9], lung function (as measured by forced expiratory volume in one second -%FEV1), and time in hospital. Postcodes were used to derive Index of Multiple Deprivation (IMD) scores for the constituent UK countries [20] and each person was allocated to a deprivation quintile on the basis of first recorded postcode. These indices combine economic, social and housing indicators measured at the census into a composite deprivation score for small areas in the UK constituent countries. Baseline covariates in the analysis were: sex; genotype coded as the number of delta F508 alleles (0, 1 or 2); and year of birth. We adjusted for disease severity on the basis of degree of impaired lung function measured by %FEV1, as this measure is strongly predictive of survival [21], and deviation from expected Body Mass Index (as measured by Body Mass Index standard deviation score (BMI SD score)). As a measure of time spent administering therapies, we included the number of intravenous (IV) therapy days in the past year in our analysis, further disaggregated into therapy days in hospital, and therapy days at home. We first fitted a model adjusted for age and the baseline covariates, which are unlikely to be in the causal path from SES to employment status. We then tested for the significance of adding disease severity measures, and service use measures, which may be in the causal path, and finally added deprivation score to the model. The logic model for the analysis is shown in Figure S1.

### Statistical Methods

Repeated measures on individuals are correlated, and this must be accommodated to obtain valid inferences. For a full description see the appendix S1. In brief, we applied advanced statistical methods that have been specifically developed for the analysis of longitudinal data, after Diggle et al, 2002 (24). Exploratory statistical analysis involved: fitting generalized additive models (GAMs) [22] to visualize the shape of associations; plotting empirical logits; and plotting stratified raw data. We then fitted generalised linear mixed models (GLMMs) to the data across the age range. These procedures model the log-odds of employment status as a linear function of the measured covariates and individual level random-effects, and adjust the standard errors of the regression parameters to take account of the correlation structure of the repeated measurements. We fitted sequential models adjusting for the covariates of interest, and estimated model parameters by maximum likelihood, using generalized likelihood ratio statistics to compare nested models, and Wald statistics to test hypotheses about model parameters [23]. These longitudinal analyses take into account drop-out due to death, and implicitly estimate the chances of employment in a hypothetical drop-out free population [24]. We present effect estimates as log-odds with confidence intervals, since odds ratios can be misinterpreted when outcomes are common [25]. To aid interpretation, we display population-averaged employment chances in the plots, by averaging individual-level fitted values over the population.

### Results

The final dataset contained 3,458 people, with 15,098 person-years of follow-up, and data collected at 15,572 annual reviews. 1940 (56%) individuals had four or more follow-up measures (median 4, interquartile range 2 to 7). The baseline characteristics of individuals at first recorded entry into the cohort are shown in table 1.

At any one time, about 50% of the UK CF population were recorded as being in full or part-time employment for all ages, but

**Table 1.** Characteristics of study population in UK CF Registry by employment status at baseline.

	Not in employment	Employed	Total
Number of adults with CF (%)	1845 (53.4)	1613 (46.6)	3458
Observations (%)	7287 (46.8)	8285 (53.2)	15572
Deprivation quintile 1 (least deprived)	295 (16)	344 (21.3)	639 (18.5)
Deprivation quintile 2	319 (17.3)	370 (22.9)	689 (19.9)
Deprivation quintile 3	357 (19.3)	354 (21.9)	711 (20.6)
Deprivation quintile 4	417 (22.6)	321 (19.9)	738 (21.3)
Deprivation quintile 5 (most deprived)	457 (24.8)	224 (13.9)	681 (19.7)
Number of F508 alleles:2 (%)	952 (51.6)	744 (46.1)	1696 (49)
Number of F508 alleles:1 (%)	632 (34.3)	616 (38.2)	1248 (36.1)
Number of F508 alleles:0 (%)	261 (14.1)	253 (15.7)	514 (14.9)
Female	856 (46.4)	672 (41.7)	1528 (44.2)
Non-white	54 (2.9)	27 (1.7)	81 (2.3)
Birth cohort 1959 – 1968	152 (8.2)	211 (13.1)	363 (10.5)
Birth cohort 1969 – 1978	409 (22.2)	573 (35.5)	982 (28.4)
Birth cohort 1979 – 1988	1203 (65.2)	780 (48.4)	1983 (57.3)
Birth cohort >1988	81 (4.4)	49 (3)	130 (3.8)
Median age at baseline (years) (IQR)	21 (20.4,24.5)	23 (20.7,29.3)	21.5 (20.5,27)
Median %FEV1 at entry (IQR)	61.8 (41.8,82.2)	68.4 (49.9,85.2)	65.3 (45.8,83.7)
%FEV1 >90 (normal)	286 (15.5)	302 (18.7)	588 (17)
%FEV1 >70 and <90 (mild)	457 (24.8)	458 (28.4)	915 (26.5)
% FEV1 >40 and <70 (moderate)	678 (36.7)	619 (38.4)	1297 (37.5)
% FEV1 <40 (severe)	424 (23)	234 (14.5)	658 (19)
Pseudomonas colonization at entry	1123 (60.9)	836 (51.8)	1959 (56.7)
Median BMI SDS at entry (IQR)	-0.6 (-1.4,0.1)	-0.4 (-1.2,0.4)	-0.5 (-1.3,0.3)
Died	223 (12.1)	115 (7.1)	338 (9.8)

doi:10.1371/journal.pone.0073322.t001

patterns differed by age, sex, and deprivation status (Figure 1). Across the entire age range, men and women with CF from the most deprived quintile were much less likely to be in employment than their counterparts in the least deprived quintile (Figure 1).

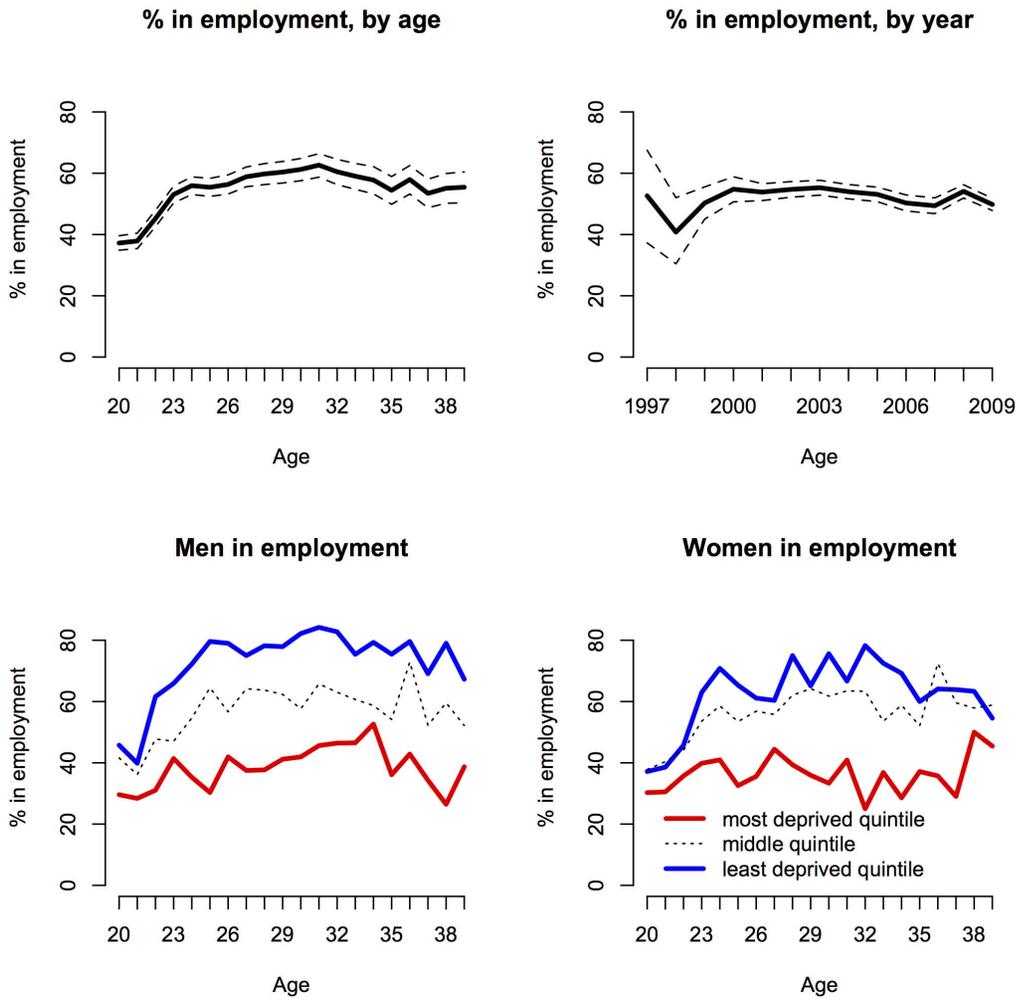
Figure 2 illustrates the modelled independent population averaged relationship between deprivation, sex, lung function, weight and time in hospital, and employment chances for people with CF in the UK, on the basis of the final interaction model (table 2, column 5). There are significant age-related effects. The general pattern was for the proportion of people in employment to increase to around age 30, and decrease subsequently. Genotype and use of home intravenous therapy were not associated with employment chances in any of the analyses, and were dropped from the final models.

Of the covariates in the model, deprivation status explained more of the variance, and there was a dose-response relationship, in that the greater the level of deprivation the lower the chances if employment (figure 2, table 2). People in the most deprived quintile were less likely to be in employment, after adjusting for disease severity, compared to their more advantaged counterparts (log-odds -2.66 95%CI -3.1 to -2.26, comparing the most to the least deprived quintile). For men with a middling level of lung function (%FEV1 of 60) at the age of 30, this equates to 67.7% employment in the least deprived quintile, compared to 50.2% in the most deprived, a difference of 17.6 percentage points (figure 2). Comparing a population with relatively good lung function (a%FEV of 80), to one with poor lung function (a%FEV1 of 30),

with all other things being equal (i.e. deprivation quintile 3, male sex), at the age of 30 there was a difference of 7.2 percentage points in employment chances (log-odds -0.63 95%CI -1.1 to -0.15 comparing poor to good lung function (figure 2).

Men were more likely than women to be in employment (log-odds 0.40 95%CI 0.16 to 0.64 in adjusted model), which corresponds to 61.7% employment in men, compared to 58.7% employment in women at age 30, for people with a%FEV1 of 60, in the middle deprivation quintile – a difference of 3 percentage points. People with better lung function were more likely to be in employment, and this followed a monotone dose response relationship. Higher BMI was associated with improved employment chances (log-odds 0.1 95%CI 0.020 to 0.188 per 1 unit increase in BMI SD score), and more days in hospital were associated with lower employment chances (log-odds -0.023 95%CI -0.027 to -0.019, per day in hospital).

Figure 3 illustrates the interactive effect of degree of disease severity (as measured by level of lung function) and social deprivation on population averaged employment chances in men and women. In the final model, a composite test for interaction between level of lung function and deprivation quintile was not significant ( $p=0.2$ ), but the contrast between the most deprived quintile and level of lung function was significant (log-odds 0.016 95% CI 0.0028 to 0.03, per unit increase in %FEV1,  $p=0.018$ ) (table 2). This suggests that having poor lung function was more damaging to the employment chances of the most disadvantaged quintile than for the least disadvantaged. At age 30, for example,



**Figure 1. Overall employment prevalence by age and year of people with CF aged 20–40 in UK CF Registry.** 95% confidence intervals (top row). Bottom row shows employment prevalence by age, stratified by deprivation quintile (most deprived quintile in red), for men and women. doi:10.1371/journal.pone.0073322.g001

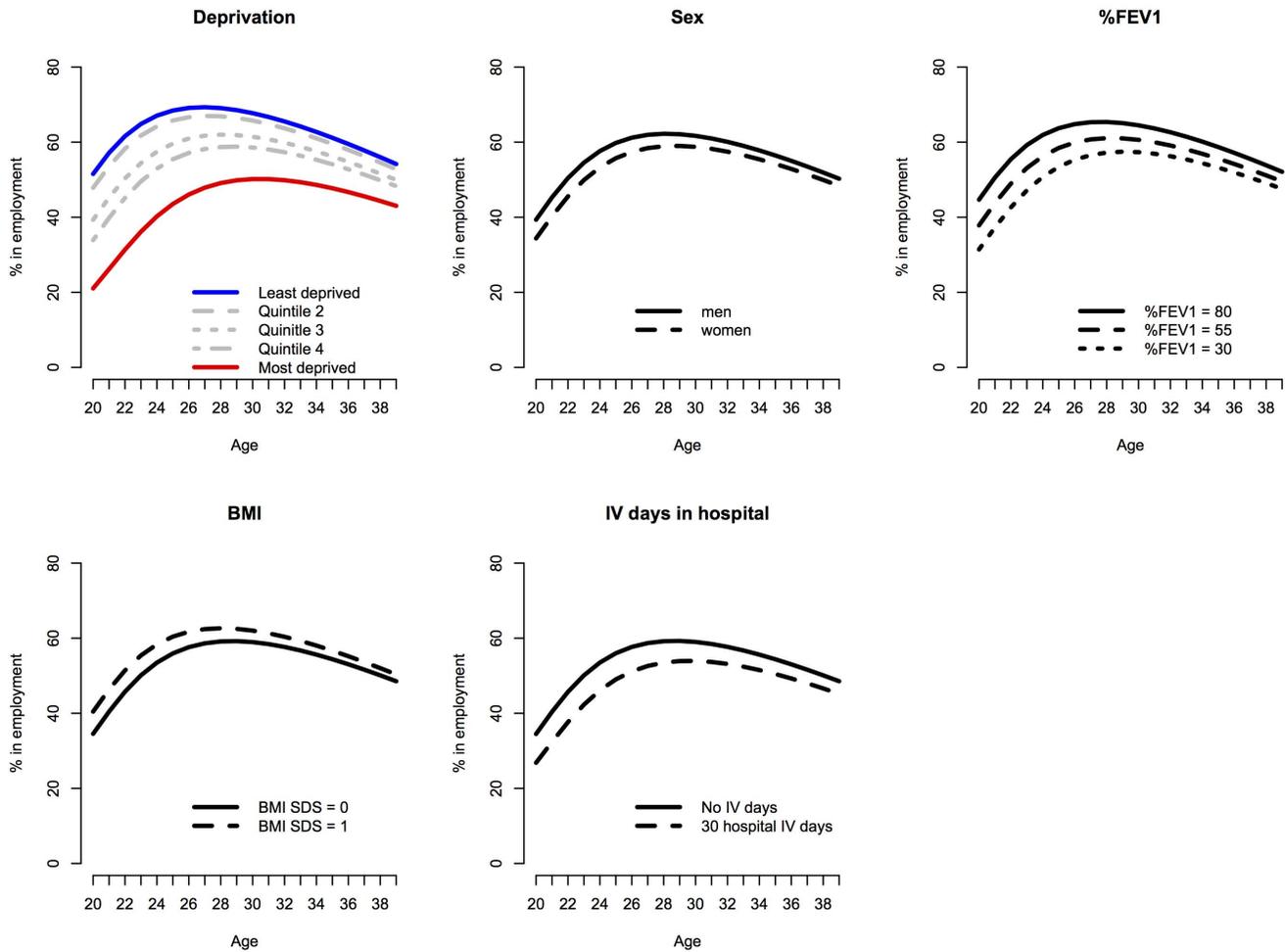
poor lung function in men in the least deprived quintile was associated with 3.1 percentage points lower employment chances than for their counterparts in the same quintile with good lung function (66.1% employment prevalence compared with 69.2%). For men in the most deprived quintile, however, there is 7.7 percentage points difference in employment chances between those with poor and good lung function (46.4% compared with 54.1%).

**Discussion**

We undertook a longitudinal registry-based study of employment status in the UK CF population, and found that lower social deprivation, male sex, higher level of lung function and BMI, and less time in hospital were associated with improved employment chances. All other things being equal, being deprived was associated with lower employment chances (17.6 percentage points lower than least deprived quintile). Having poor lung function was also associated with lower employment chances (7.2 percentage points lower in people with poor versus good lung function). When people with CF have a double burden of high deprivation and poor lung function, however, the impact on employment chances is magnified. In other words,, deprivation

appears to modify the effect of lung function on employment chances - poor lung function is more harmful to employment chances in the most deprived areas, compared to the least. As people are living longer, healthier lives with CF, it is more important than ever for strategies to promote employment to focus on the broader societal barriers to engagement in the workforce for people with CF, rather than taking a narrow ‘impairment’ focus solely on the impact of disease severity on employment chances, as critiqued by Edwards et al [8].

Key strengths of this study include the population-wide coverage of the UK CF registry, the high quality of the data, the longitudinal analysis, and the theoretical approach that responds to previous criticisms of the illness-focussed approach to understanding employment outcomes in CF. Our findings are particularly relevant to the UK population, but could be cautiously generalised to other high-income countries. There are limitations: it relies on retrospective, routinely collected data and although we used a standard, fine-grained measure of deprivation of area of residence, employed widely in epidemiological studies in the UK as a measure of socio-economic status (SES) [9,26,27], it was not possible to separate effects of socioeconomic circumstances operating at the individual and area level. There is thus the possibility of ecological bias, and this limits the possibilities to



**Figure 2. Longitudinal employment trajectory versus age of people with CF in UK CF Registry, by deprivation quintile, sex, %FEV1, BMI SD score and days in hospital.** The lines show the final modelled longitudinal trajectories from the final interaction model (table 2, column 5), contrasting the adjusted effects of deprivation, sex, %FEV1, BMI SD score and days in hospital. These effects are plotted at the reference levels for the other covariates in the analysis. doi:10.1371/journal.pone.0073322.g002

disentangle the mechanisms by which socioeconomic disadvantage operates. However, this possibility is minimized using the established Index of Multiple Deprivation methodology in the UK [9]. Secondly, we only had valid postcodes on 90% of the sample, though our sample size was large, with no pronounced gradient in the proportion of patients with valid postcodes by deprivation quintile i.e. missing postcode information was unrelated to deprivation. Finally the analysis did not include data for people aged over 40 years, because less than 5% of the annual clinical reviews occurred in patients over 40 years, so, although including these data would extend the age range for the analysis considerably, there would be small numbers in the over-40 age-range.

A recent review concluded that further research on CF and employment is necessary to improve occupational outcomes [7]. The systematic review identified nine studies that have looked at the relationship between CF and employment status, all of which were small, and based at one or two care centres only. Six studies reported employment rates, all of around 50%. This systematic review did not include the largest study to date, by Walters et al [28], which was a cross-sectional questionnaire survey of 1052 adults over 16 years of age with CF in the UK in 1990. Walters et

al found that 55% of responders were working, whilst of those not employed, half gave ill health as the reason [28]. In our study we find that the annual employment rate among CF adults in the UK appears to have remained unchanged between 50% and 55% over the last decade. Five previous studies in the systematic review analysed the association between disease severity, as measured by lung function, and employment status, but with no analysis by SES. The results were mixed: three studies conducted in the US [29], Canada [30] and Australia [31] concluded that lung function was not related to employment status, and two from the US [32] and Belgium [33] suggested that it was related. The studies by Burker et al. [29], Gillen et al. [32], and Hogg et al. [31] suggested that frequency of hospital admissions, demographic variables, mental health, and education level, may also influence employment status.

Our study suggests that disease severity and time in hospital influence employment chances in the UK CF population, but these effects are not as large as one might have predicted. It is evident that people with significant respiratory impairment continue to work, and disease severity alone does not predict employment outcomes. Furthermore, there is a large amount of variability between individuals with similar characteristics (large

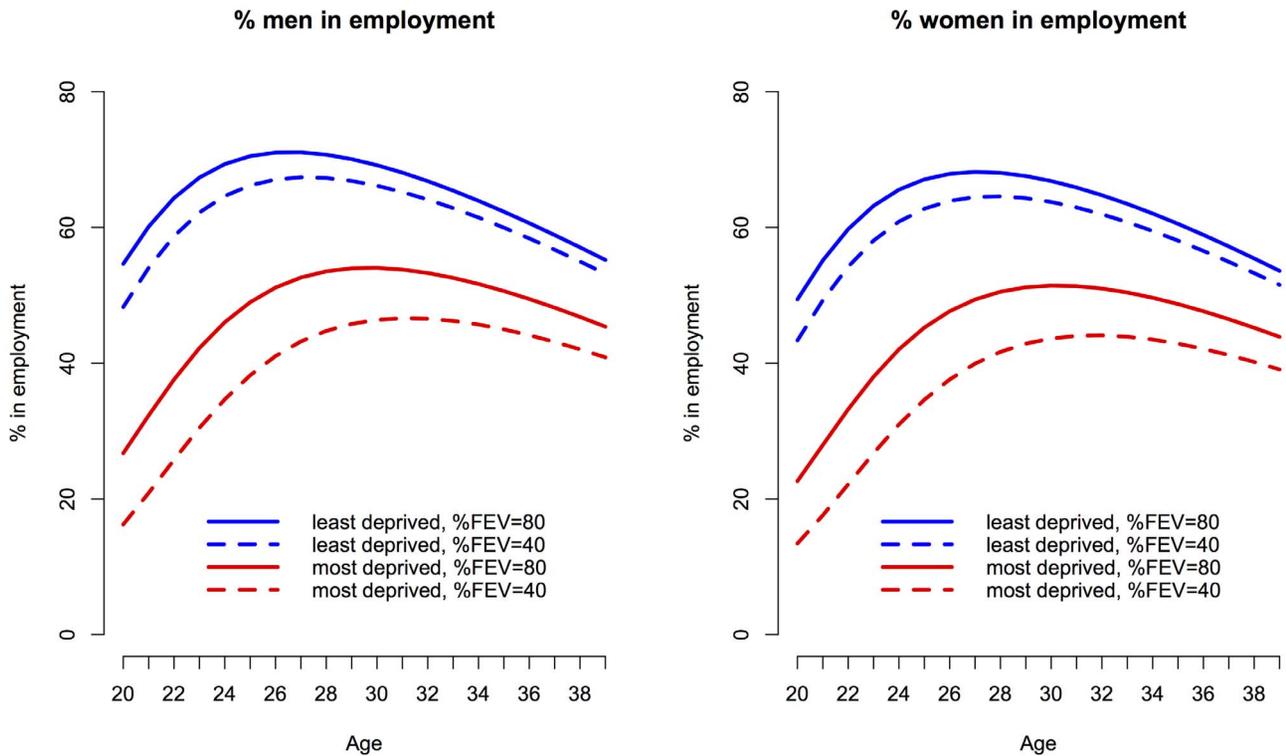
**Table 2.** Log odds for the final nested generalised mixed effects models (GLMMs) for the effects of disease severity, time in hospital and level of deprivation on employment chances.

	<i>Baseline</i>	<i>Baseline +severity</i>	<i>Baseline +Severity +Time in hospital</i>	<i>Baseline +Severity +Time in hospital +Deprivation</i>	<i>Baseline +Severity +Time in hospital +Deprivation x%FEV1</i>
Constant	0.663*** (0.105)	0.761*** (0.106)	0.997*** (0.105)	2.047*** (0.172)	2.058*** (0.171)
age	0.128*** (0.016)	0.143*** (0.016)	0.144*** (0.016)	0.146*** (0.016)	0.146*** (0.016)
age <sup>2</sup>	-0.022*** (0.002)	-0.024*** (0.002)	-0.023*** (0.002)	-0.023*** (0.002)	-0.023*** (0.002)
Birthyear	-0.057*** (0.014)	-0.059*** (0.014)	-0.046** (0.014)	-0.039** (0.014)	-0.039** (0.014)
Male/Female	0.446*** (0.129)	0.485*** (0.128)	0.443*** (0.125)	0.410** (0.125)	0.401** (0.125)
Random intercept SD	(3.016)	(2.893)	(2.779)	(2.640)	(2.632)
Random slope SD	(0.450)	(0.451)	(0.438)	(0.442)	(0.441)
%FEV1		0.026*** (0.002)	0.020*** (0.002)	0.021*** (0.002)	0.013** (0.005)
BMI SDS score		0.163*** (0.043)	0.121** (0.043)	0.106* (0.043)	0.104* (0.043)
Hospital IV days			-0.024*** (0.002)	-0.023*** (0.002)	-0.023*** (0.002)
Deprivation quintile 2/1				-0.270 (0.202)	-0.279 (0.202)
Deprivation quintile 3/1				-0.967*** (0.200)	-0.976*** (0.200)
Deprivation quintile 4/1				-1.422*** (0.198)	-1.427*** (0.198)
Deprivation quintile 5/1				-2.650*** (0.207)	-2.663*** (0.207)
Deprivation quintile 2/1 x%FEV1					0.010 (0.007)
Deprivation quintile 3/1 x%FEV1					0.010 (0.006)
Deprivation quintile 4/1 x%FEV1					0.006 (0.006)
Deprivation quintile 5/1 x%FEV1					0.016* (0.007)
Log-likelihood	-7885	-7720	-7645	-7548	-7545
N	15572	15430	15430	15430	15430
Groups	3458	3451	3451	3451	3451
Baseline variance explained (%)	-	7.9	15.1	23.3	23.8

\* $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .  
Standard errors in parentheses.  
doi:10.1371/journal.pone.0073322.t002

random effects), which suggests that there are other important factors related to employment status that we have been unable to account for in our study. Highest educational attainment is one such factor, but these data were available only for 60% of the

individuals in the UK CF Registry. Adjusting for educational attainment did not change any of the substantive effects in our analysis, though it did reduce the random effects variance further (table S1). Other studies have also suggested that individual



**Figure 3. Longitudinal employment trajectory versus age, demonstrating the interaction between deprivation and lung function.** The lines show the final modelled longitudinal trajectories from the interaction model (table 2, column 5), contrasting the adjusted effects of deprivation, and %FEV1.  
doi:10.1371/journal.pone.0073322.g003

psychological factors and education status are correlated with employment status in CF [29].

These previous studies on employment chances in people with CF tend to portray CF as a ‘serious illness’, which causes employment problems. In contrast, Edward et al explored the employment experiences of adults with CF from a social model perspective. They demonstrated barriers to employment that were similar to those experienced by other disabled people, as well as barriers related to the ‘impairment effects’ of CF, and concluded that adults with CF have valuable perspectives to contribute to social model analysis and the development of employment-related policy solutions [8]. Our results not only corroborate, but also extend these observations, by demonstrating the interaction between disease severity related factors, and deprivation.

Our findings add to the extensive literature on the inter-relationship between chronic illness, socioeconomic status, and employment opportunities [34]. In the UK in 2005 the age standardized employment rate for people of working age (25–59) was 80% in healthy women, compared to 50% in those with limiting long-standing illness (LLSI), and 93% compared to 59% in men [34]. Furthermore there was a striking social gradient for those with LLSI - the prevalence of employment was 66% in highly educated women with LLSI compared to 18% in those with low education, and in men 72% compared to 30% [35]. In our study, the unadjusted prevalence of employment was around 60% in the most affluent quintile, compared to 30% in the most deprived quintile in women, and 70% versus 30% in men (figure 1).

Low employment in people with CF is a serious concern. Despite there being no difference in incidence of CF by socioeconomic status (SES), there are important differences in

outcomes such as growth and lung function, and ultimately survival, in people with CF by SES in the UK and US [9,12,15]. Being out of work increases the risk of poverty and social exclusion, and is likely to further damage the health of the most disadvantaged people with CF. In this study we have demonstrated *differential social consequences of illness* in the context of CF, by which people with the double burden of chronic illness and low SES are more likely to be excluded from the labour market. We speculate that this may be an important pathway for the amplification of health inequalities in CF, whereby disadvantage builds on disadvantage. It is of particular concern that the most disadvantaged women have the poorest employment chances, since female sex is also an important risk factor for poor survival in CF [15].

In conclusion, this study has identified important longitudinal inequalities in employment outcomes in people with CF in the UK. Future studies should focus on policy interventions that would help overcome the extra burden of adverse consequences of CF faced by patients living in disadvantaged circumstances.

**Supporting Information**

**Figure S1** Logic model to inform analysis of employment status. (PDF)

**Table S1** Log odds for the final generalised mixed effects models (GLMMs), with added educational variable. (DOCX)

**Appendix S1** Supplementary text. (DOCX)

## Acknowledgments

We wish to thank the UK Cystic Fibrosis Trust for access to the UK CF Registry and all the Centre Directors for the input of data to the registry. Elaine Gunn assisted with access to the data.

## References

1. Cystic Fibrosis Trust (2011) UK CF Registry annual data report 2010. <http://www.cftrust.org.uk/aboutcf/publications/cfregistryreports/> (accessed 28th September 2012).
2. Dodge JA, Lewis PA, Stanton M, Wilsher J (2007) Cystic fibrosis mortality and survival in the UK: 1947–2003. *European Respiratory Journal* 29: 522–526.
3. Marmot MG, Allen J, Goldblatt P, Boyce T, McNeish D, et al. (2010) Fair society, healthy lives: Strategic review of health inequalities in England post-2010. *The Marmot Review*: London UK. 2010.
4. Bamba C, Gibson M, Sowden AJ, Wright K, Whitehead M, et al. (2009) Working for health? Evidence from systematic reviews on the effects on health and health inequalities of organisational changes to the psychosocial work environment. *Prev Med* 48: 454–461.
5. Bamba C, Gibson M, Sowden A, Wright K, Whitehead M, et al. (2010) Tackling the wider social determinants of health and health inequalities: evidence from systematic reviews. *J Epidemiol Community Health* 64: 284–291.
6. Sawicki GS, Sellers DE, Robinson WM (2009) High treatment burden in adults with cystic fibrosis: challenges to disease self-management. *J Cyst Fibros* 8: 91–96.
7. Saldana PS, Pomeranz JL (2012) Cystic fibrosis and the workplace: a review of the literature. *Work* 42: 185–193.
8. Edwards J, Boxall K (2010) Adults with cystic fibrosis and barriers to employment. *Disability & Society* 25: 441–453.
9. Taylor-Robinson D, Smyth RL, Diggle P, Whitehead M (2013) The effect of social deprivation on clinical outcomes and the use of treatments in the UK cystic fibrosis population: A longitudinal study *Lancet Respiratory Medicine* 1: 121–128.
10. Taylor-Robinson D, Schechter MS (2011) Health inequalities and cystic fibrosis. *BMJ* 343: d4818.
11. Diderichsen F, Evans T, Whitehead M (2001) *The social origins of disparities in health. Challenging inequities in health.* New York: Oxford University Press.
12. Schechter MS, Shelton BJ, Margolis PA, Fitzsimmons SC (2001) The association of socioeconomic status with outcomes in cystic fibrosis patients in the United States. *Am J Respir Crit Care Med* 163: 1331–1337.
13. O'Connor GT, Quinton HB, Kneeland T, Kahn R, Lever T, et al. (2003) Median household income and mortality rate in cystic fibrosis. *Pediatrics* 111: e333–339.
14. Britton JR (1989) Effects of social class, sex, and region of residence on age at death from cystic fibrosis. *BMJ* 298: 483–487.
15. Barr HL, SA, Fogarty AW (2011) The association between socioeconomic status and gender with median age at death from cystic fibrosis in England and Wales: 1959 to 2008. *BMJ*.
16. Ben-Shlomo Y, Kuh D (2002) A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *International Journal of Epidemiology* 31: 285.
17. Cystic Fibrosis Trust (2011) UK CF Registry.

## Author Contributions

Conceived and designed the experiments: DTR PD MMW RLS. Performed the experiments: DTR PD. Analyzed the data: DTR PD. Contributed reagents/materials/analysis tools: DTR PD MMW RLS. Wrote the paper: DTR PD MMW RLS.

18. Adler AI, Shine BS, Chamnan P, Haworth CS, Bilton D (2008) Genetic determinants and epidemiology of cystic fibrosis-related diabetes: results from a British cohort of children and adults. *Diabetes Care* 31: 1789–1794.
19. Mehta G, Sims EJ, Culross F, McCormick JD, Mehta A (2004) Potential benefits of the UK Cystic Fibrosis Database. *J R Soc Med* 97 Suppl 44: 60–71.
20. GeoConvert (2011) <http://geoconvert.mimas.ac.uk/> (accessed 29th July 2011).
21. Rosenfeld M, Coates A, Corey M, Goss CH, Howard MB, et al. (2005) Task Force to Evaluate Choice of Spirometric Reference Equations for the National Patient Registry: Summary and Recommendations. Cystic Fibrosis Foundation Registry Committee; Oct 1, 2005.
22. Hastie TJ, Tibshirani RJ (1990) *Generalized additive models*: Chapman & Hall/CRC.
23. Laird NM, Ware JH (1982) Random-effects models for longitudinal data. *Biometrics* 38: 963–974.
24. Diggle P, Heagerty P, Liang K-Y, Zeger SL (2002) *Analysis of Longitudinal Data* (second edition). Oxford: Oxford University Press.
25. Grimes DA, Schulz KF (2008) Making sense of odds and odds ratios. *Obstet Gynecol* 111: 423–426.
26. Semple MG, Taylor-Robinson DC, Lane S, Smyth RL (2011) Household tobacco smoke and admission weight predict severe bronchiolitis in infants independent of deprivation: prospective cohort study. *PLoS One* 6: e22425.
27. Bergen H, Hawton K, Waters K, Ness J, Cooper J, et al. (2012) Premature death after self-harm: a multicentre cohort study. *Lancet* 380: 1568–1574.
28. Walters S, Britton J, Hodson ME (1993) Demographic and social characteristics of adults with cystic fibrosis in the United Kingdom. *BMJ* 306: 549–552.
29. Burker EJ, Sedway J, Carone S (2004) Psychological and educational factors: better predictors of work status than FEV1 in adults with cystic fibrosis. *Pediatr Pulmonol* 38: 413–418.
30. Frangolias DD, Holloway CL, Vedal S, Wilcox PG (2003) Role of exercise and lung function in predicting work status in cystic fibrosis. *Am J Respir Crit Care Med* 167: 150–157.
31. Hogg M, Braithwaite M, Bailey M, Kotsimbos T, Wilson JW (2007) Work disability in adults with cystic fibrosis and its relationship to quality of life. *J Cyst Fibros* 6: 223–227.
32. Gillen M, Lallas D, Brown C, Yelin E, Blanc P (1995) Work disability in adults with cystic fibrosis. *Am J Respir Crit Care Med* 152: 153–156.
33. Havermans T, Colpaert K, Vanharen L, Dupont LJ (2009) Health related quality of life in cystic fibrosis: To work or not to work? *J Cyst Fibros* 8: 218–223.
34. Holland P, Nysten L, Thielen K, van der Wel KA, Chen WH, et al. (2011) How do macro-level contexts and policies affect the employment chances of chronically ill and disabled people? Part II: The impact of active and passive labor market policies. *Int J Health Serv* 41: 415–430.
35. Holland P, Burstrom B, Whitehead M, Diderichsen F, Dahl E, et al. (2011) How do macro-level contexts and policies affect the employment chances of chronically ill and disabled people? Part I: The impact of recession and deindustrialization. *Int J Health Serv* 41: 395–413.

## ORIGINAL ARTICLE

# Understanding the natural progression in %FEV<sub>1</sub> decline in patients with cystic fibrosis: a longitudinal study

David Taylor-Robinson,<sup>1</sup> Margaret Whitehead,<sup>1,2</sup> Finn Diderichsen,<sup>2</sup> Hanne Vebert Olesen,<sup>3</sup> Tania Pressler,<sup>4</sup> Rosalind L Smyth,<sup>5</sup> Peter Diggle<sup>6</sup>

► Additional materials are published online only. To view these files please visit the journal online (<http://thorax.bmj.com/content/early/recent>).

<sup>1</sup>Department of Public Health and Policy, University of Liverpool, Liverpool, UK

<sup>2</sup>Department of Social Medicine, University of Copenhagen, Copenhagen, Denmark

<sup>3</sup>Cystic Fibrosis Center, Aarhus University Hospital, Aarhus, Denmark

<sup>4</sup>Cystic Fibrosis Center, Rigshospitalet, Copenhagen, Denmark

<sup>5</sup>Division of Child Health, University of Liverpool, Liverpool, UK

<sup>6</sup>School of Health and Medicine, Lancaster University, Lancaster, UK

**Correspondence to**

Dr David Taylor-Robinson, MRC Population Health Scientist, Department of Public Health and Policy, Whelan Building, University of Liverpool, Liverpool L69 3GB, UK; [dctr@liv.ac.uk](mailto:dctr@liv.ac.uk)

RLS and PJD are joint senior authors.

Received 15 August 2011  
Accepted 26 February 2012



This paper is freely available online under the BMJ Journals unlocked scheme, see <http://thorax.bmj.com/site/about/unlocked.xhtml>

**ABSTRACT**

**Background** Forced expiratory volume in 1 s as a percentage of predicted (%FEV<sub>1</sub>) is a key outcome in cystic fibrosis (CF) and other lung diseases. As people with CF survive for longer periods, new methods are required to understand the way %FEV<sub>1</sub> changes over time. An up to date approach for longitudinal modelling of %FEV<sub>1</sub> is presented and applied to a unique CF dataset to demonstrate its utility at the clinical and population level.

**Methods and findings** The Danish CF register contains 70 448 %FEV<sub>1</sub> measures on 479 patients seen monthly between 1969 and 2010. The variability in the data is partitioned into three components (between patient, within patient and measurement error) using the empirical variogram. Then a linear mixed effects model is developed to explore factors influencing %FEV<sub>1</sub> in this population. Lung function measures are correlated for over 15 years. A baseline %FEV<sub>1</sub> value explains 63% of the variability in %FEV<sub>1</sub> at 1 year, 40% at 3 years, and about 30% at 5 years. The model output smooths out the short-term variability in %FEV<sub>1</sub> (SD 6.3%), aiding clinical interpretation of changes in %FEV<sub>1</sub>. At the population level significant effects of birth cohort, pancreatic status and *Pseudomonas aeruginosa* infection status on %FEV<sub>1</sub> are shown over time.

**Conclusions** This approach provides a more realistic estimate of the %FEV<sub>1</sub> trajectory of people with chronic lung disease by acknowledging the imprecision in individual measurements and the correlation structure of repeated measurements on the same individual over time. This method has applications for clinicians in assessing prognosis and the need for treatment intensification, and for use in clinical trials.

**INTRODUCTION**

Understanding the long-term natural history of changes in lung function in people with lung diseases is a research priority.<sup>1</sup> In order to do this, objective measures of disease progression are necessary. The per cent predicted forced expiratory volume in 1 s (%FEV<sub>1</sub>) is commonly used to monitor lung function, and to describe disease severity in cystic fibrosis (CF)<sup>2</sup> and chronic obstructive pulmonary disease (COPD).<sup>3</sup> %FEV<sub>1</sub> is used to inform clinical decisions about changing or intensifying treatment, and as an outcome measure in clinical studies.<sup>4–6</sup> Furthermore %FEV<sub>1</sub> has been shown to be related to survival in CF. Kerem *et al*'s

**Key messages****What is the key question?**

► Now that people with cystic fibrosis are living much longer, how can we optimally describe the changes in forced expiratory volume in 1 s as a percentage of predicted (%FEV<sub>1</sub>) over time in a way that is useful for clinicians at the individual and the population level?

**What is the bottom line?**

► We describe a novel modelling approach for analysing changes in %FEV<sub>1</sub> over time that can be applied at the individual level to interpret the clinical significance of sudden changes in %FEV<sub>1</sub>, and at the population level to quantify the effect of factors such as *Pseudomonas aeruginosa* acquisition.

**Why read on?**

► Lung function measures are correlated for over 15 years, and a baseline %FEV<sub>1</sub> value explains 63% of the variability in %FEV<sub>1</sub> at 1 year, 40% at 3 years and about 30% at 5 years.

study in 1992 demonstrated that patients with a %FEV<sub>1</sub> <30 had a 2-year mortality over 50%,<sup>7</sup> though a more recent study shows that survival rates at low levels of lung function have improved in subsequent cohorts.<sup>8</sup>

Interpreting the significance of changes in %FEV<sub>1</sub> in CF to inform patient management and to counsel patients regarding prognosis requires an understanding of the inherent variability of %FEV<sub>1</sub> measures within individuals, to determine what constitutes a clinically significant deterioration in %FEV<sub>1</sub>, rather than a change due to measurement error, or recoverable day-to-day fluctuation in lung function.<sup>9 10</sup> Furthermore, this variability needs to be understood to make valid inferences about the association between covariates and %FEV<sub>1</sub> in observational studies.

As survival in CF improves with successive cohorts, there are many more people surviving into late adulthood. An implication of this, coupled with the availability of long-term follow-up data in CF registers, is that up to date methods should be adopted to interpret the long-term dynamics of lung function in CF. Statistical techniques for

## Cystic fibrosis

longitudinal data analysis have been the subject of much methodological development over the past 20 years, and the random intercept and slope model has become a popular analysis framework.<sup>4 5 11–14</sup> While this is often appropriate for relatively short follow-up periods, there are theoretical reasons to suggest that this approach makes assumptions that will lead to incorrect inferences if applied over longer follow-up periods. One central assumption is that the variability in %FEV<sub>1</sub> increases as a quadratic function over time (in proportion to time squared), which leads to estimates that diverge unrealistically over longer time periods. Methods for undertaking these analyses over longer time periods have been described,<sup>15</sup> but have not been commonly applied.

In this study we analyse a unique population-level dataset of people with CF that includes longitudinal %FEV<sub>1</sub> measures taken monthly for up to 30 years. We apply these methods to develop a general model for %FEV<sub>1</sub> decline that goes beyond the popular random-intercept and slope approach, and explicitly describes the variability in %FEV<sub>1</sub> within individuals over time. We show how this could be applied clinically to help interpret the significance of changes in lung function, and at a population level to explore the association of covariates (eg, *Pseudomonas aeruginosa* acquisition) with %FEV<sub>1</sub> decline.

## METHODS

### Subjects

All patients aged over 5 years whose %FEV<sub>1</sub> data were entered on the Danish CF database between 1969 and 2010 were eligible. Post-transplant data from patients who had received a lung transplant were excluded. Patients attending the two Danish CF centres (Copenhagen and Aarhus) are seen routinely every month in the outpatient clinic for evaluation of clinical status, pulmonary function and microbiology of lower respiratory tract secretions. It is estimated that coverage of people with CF resident in Denmark is almost complete from 1990 when CF care was centralised. This coverage and the unparalleled frequency of measurement make this a unique dataset for epidemiological research. The study was approved by the Danish Data Inspectorate (Datatilsynet).

### Lung function testing

The primary outcome for this analysis was %FEV<sub>1</sub>. Pulmonary function tests were performed according to international recommendations,<sup>16</sup> measuring FEV<sub>1</sub>, expressed as a percentage of predicted values for sex and height using reference equations from Wang or Hankinson.<sup>17 18</sup>

## Covariates

Covariates in the analysis were age, sex, genotype coded as the number of Delta F508 alleles (0, 1 or 2), onset of chronic *Pseudomonas* infection (coded 0 or 1 as a time-varying covariate), pancreatic insufficiency determined on the basis of pancreatic enzyme usage (coded 0 or 1 as a baseline covariate), birth cohort (six 10-year cohorts starting at 1948), and CF-related diabetes (CFRD) diagnosed using the WHO criteria (coded 0 or 1 as a time-varying covariate).

## Statistical analysis

A detailed explanation is given in the online appendix. Repeated %FEV<sub>1</sub> measures on individuals are correlated, and this must be accommodated to obtain valid inferences. We used a linear mixed effects model with longitudinally structured correlation,<sup>15 19</sup> and contrasted our approach with the widely used random intercept and slope model.<sup>20</sup> We modelled random variation in %FEV<sub>1</sub> over time for an individual subject so that the strength of the correlation of the random variation between two values depends on the corresponding time separation. The model decomposed the overall random variation in the data into three components: between subjects, between times within subjects, and measurement error.

First, we fit a provisional model for the mean response by ordinary least squares and used the empirical variogram of the residuals (see figure E1 in the online appendix) to provide initial estimates for the three components of variation, and for the shape of the correlation function of the between-times-within-subjects component. We then re-estimated all of the model parameters by maximum likelihood estimation, and used generalised likelihood ratio statistics to compare nested models, and Wald statistics to test hypotheses about model parameters. We assessed associations between single or multiple covariates and the population mean %FEV<sub>1</sub> over time, and explored alternatives to a linear function for the population-averaged time trend.

## RESULTS

### Population characteristics

The dataset contained 70 448 lung function measures on 479 patients seen between 1969 and 2010 in Denmark (table 1). The median number of %FEV<sub>1</sub> measures per person was 101 (range 2–597). The median follow-up period was 10.5 years (range 0.1–31.5), with a total of 6500 person-years of follow-up. Forty-two patients were followed up for more than 30 years (see also figures E2 and E3 in the online appendix).

**Table 1** Baseline characteristics of the Danish cystic fibrosis (CF) population

	Birth cohort						Total
	≥1948	≥1958	≥1968	≥1978	≥1988	≥1998	
N (%)	7 (1.5)	42 (8.8)	110 (23)	105 (21.9)	141 (29.4)	74 (15.4)	479 (100)
Women	1 (14.3)	19 (45.2)	48 (43.6)	52 (49.5)	74 (52.5)	42 (56.8)	236 (49.3)
No. Delta F508 = 0	0 (0)	0 (0)	1 (0.9)	4 (3.8)	5 (3.5)	5 (6.8)	15 (3.1)
No. Delta F508 = 1	2 (28.6)	14 (33.3)	26 (23.6)	24 (22.9)	42 (29.8)	19 (25.7)	127 (26.5)
No. Delta F508 = 2	5 (71.4)	28 (66.7)	83 (75.5)	77 (73.3)	94 (66.7)	50 (67.6)	337 (70.4)
Developed chronic <i>Pseudomonas</i>	6 (85.7)	31 (73.8)	84 (76.4)	55 (52.4)	20 (14.2)	5 (6.8)	201 (42)
Missing infection information	0 (0)	5 (11.9)	2 (1.8)	2 (1.9)	1 (0.7)	0 (0)	10 (2.1)
Pancreatic insufficient	7 (100)	42 (100)	105 (95.5)	99 (94.3)	133 (94.3)	73 (98.6)	459 (95.8)
Copenhagen	7 (100)	38 (90.5)	83 (75.5)	72 (68.6)	79 (56)	50 (67.6)	329 (68.7)
Alive	4 (57.1)	27 (64.3)	79 (71.8)	77 (73.3)	132 (93.6)	74 (100)	393 (82)
Developed CFRD	3 (42.9)	21 (50)	41 (37.3)	31 (29.5)	22 (15.6)	1 (1.4)	119 (24.8)

CFRD, cystic fibrosis related diabetes.

### Limitations of random intercept and slope model

The high degree of short-term and long-term variation in predicted %FEV<sub>1</sub> is illustrated in figure 1. The standard random intercept and slope model approach is illustrated over long and short follow-up periods in figure 1A,C. This approach assumes that any deviation of an individual's trajectory from the population mean is linear in time over the whole of the follow-up period apart from independent random errors. One can see that this assumption is reasonable over short time periods, as illustrated by the fit of the shorter dotted-line segments (figure 1A, C), but over longer time periods the individual data traces diverge unrealistically from their fitted linear mean trajectories (long solid line). Our proposed model produces a much closer fit to the data (figure 1B,D), and one that better reflects the relative magnitude of the three estimated components of variation in %FEV<sub>1</sub> over time.

### Quantifying the variability in %FEV<sub>1</sub> over time

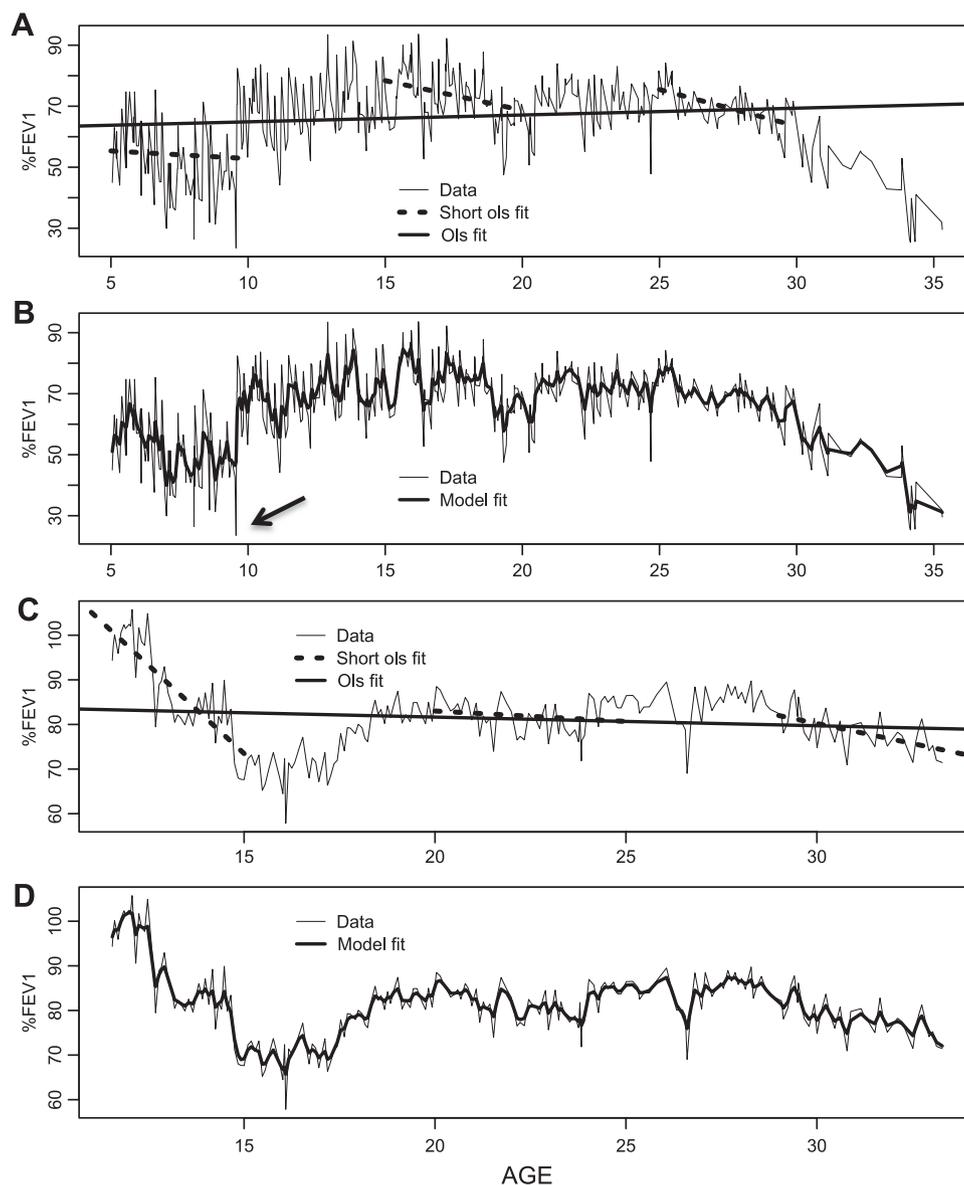
The empirical variogram quantifies the variability in the dataset (figure 2A). The intercept at time zero represents measurement error because there can be no true within-person variation at

a time lag of zero. Of the total variance in the Danish dataset, about half is due to systematic differences between patients (eg, genotype, sex or pancreatic status), two-fifths is within patients, representing change over time (disease progression), and one-tenth is 'measurement error'. In practice, this last component represents the combined effects of technical errors, and physiological variability occurring at time intervals less than the monthly interval of measurement, for example, day-to-day variability. This error variance equates to an average SD of 6.3% for repeated measures on the same individual at short time intervals. Figure 2B shows the proportion of the within-person variability in %FEV<sub>1</sub> at follow-up time (t), which can be explained by their %FEV<sub>1</sub> value at baseline. For example, about 50% of the within-patient variability at t=2.5 years is explained by the baseline measurement, and about 30% at t=5 years. Overall, the dependence on baseline measures gradually decays and is negligible at 15 years.

### Clinical utility of our proposed model

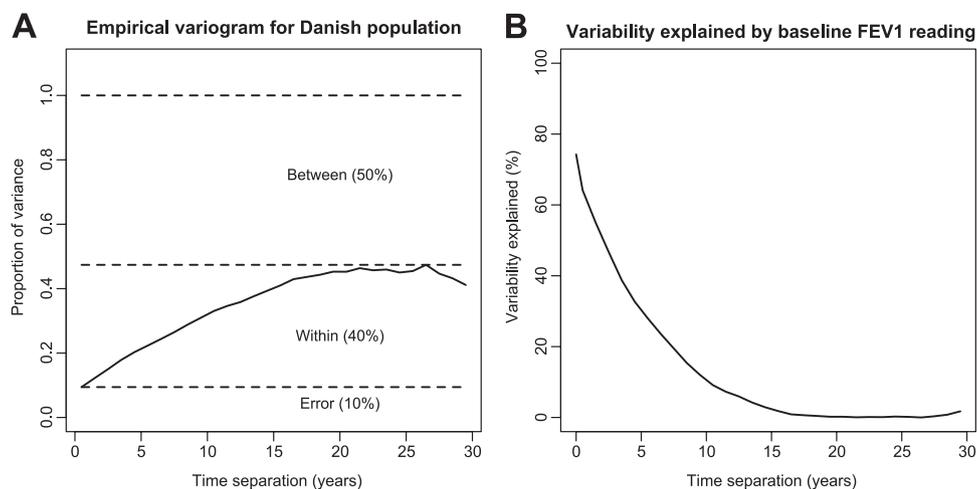
The model can be used to guide interpretation of sudden changes in lung function. Consider seeing the person in figure 1B at

**Figure 1** Comparison of conventional random intercept and slope model over short and long follow-up periods, versus our proposed Gaussian process model. (A) Data for a single individual, illustrating that a linear trend fits reasonably well over short time periods, but gives a very poor fit to this individual's complete data; linear trends are fitted by ordinary least squares. (B) The same data with the fitted trajectory of the stationary Gaussian process model. The smoothed fitted trace is a better representation of the 'true' underlying lung function, and could be used in real time to guide the interpretation of sudden changes in lung function. For instance, the sudden drop to under 30% indicated by the arrow is not mirrored in the model trace, suggesting that this may be recoverable random fluctuation. (C, D) Corresponding plots for a second individual. %FEV<sub>1</sub>, forced expiratory volume in 1 s as a percentage of predicted.



## Cystic fibrosis

**Figure 2** Quantifying the variability in forced expiratory volume in 1 s as a percentage of predicted (%FEV<sub>1</sub>) with the variogram approach. (A) Scaled empirical variogram for the Danish data. The solid line (variogram function) represents the variance of the difference between residual errors within individuals at time lags from 0 to 30 years. The variogram function increases up to about 15 years, corresponding to a decreasing correlation between paired lung function measures with increasing time separation. The variogram partitions the variability in the data into three components: within person, between person, and error. (B) Proportion of variability in an individual's %FEV<sub>1</sub> at follow-up time *t* that is explained by their %FEV<sub>1</sub> at baseline. This shows that the variogram can predict 63% of the variability from the population average at 1 year, which decreases to around 60%, 40%, 30% and 10% at 2, 3, 5 and 10 years respectively.



around age 9 (as indicated by the arrow in the figure), when her lung function has dropped to below 30%. On the basis of this one-off measurement, one might be quite guarded in terms of prognosis. However, our modelled trace (thick black line in figure 1B) suggests that her underlying lung function is changing less dramatically, with a modelled %FEV<sub>1</sub> of around 50%. We suggest that this estimate provides a more realistic assessment of underlying lung function by smoothing out the short-term variability. This could be a useful adjunct to clinical decision-making. As well as providing information about the significance of a sudden change in lung function, figure 2B also quantifies the predictive value of a contemporary %FEV<sub>1</sub> measure. In terms of counselling patients, this means that a higher %FEV<sub>1</sub> today is associated with a higher %FEV<sub>1</sub> at subsequent time points, but the predictive value deteriorates over time as illustrated in the figure.

### Effect of covariates on lung function in the Danish population

We explored the effect of covariates that have been associated with %FEV<sub>1</sub> in previous studies to demonstrate how this model can be used to answer questions at the population level (see table E1 online appendix for univariate associations).<sup>4</sup> There was

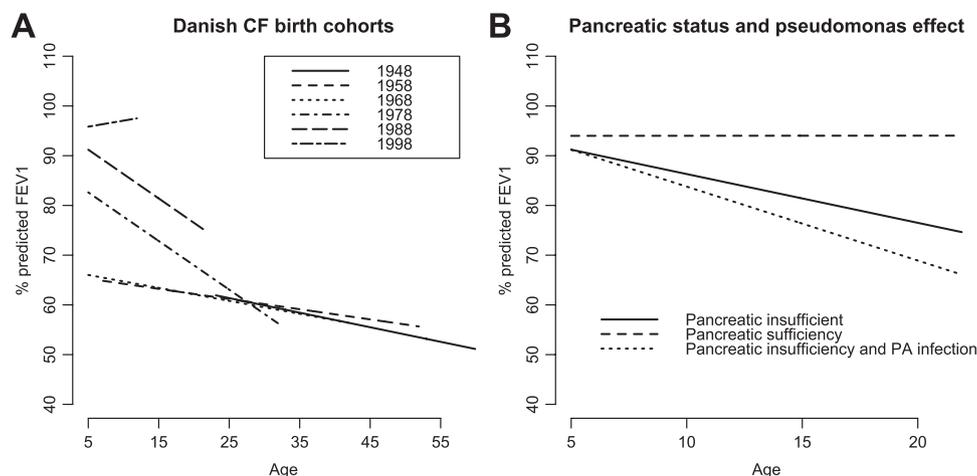
no evidence to suggest that covariate effects were nonlinear (see figure E4 in online appendix). The final model included age, *Pseudomonas* status, pancreatic status, cohort and CFRD (table 2). Note that the estimated covariate effects in table 2 are population-averaged effects, that is, they describe average values of %FEV<sub>1</sub> for sub-populations of individuals sharing the same explanatory characteristics, rather than for any one individual. The most prominent effects are associated with birth cohort, pancreatic function and the onset of *Pseudomonas* infection (figure 3). There is clear separation between the three most recent birth cohorts, with a successive increase in the intercept term at age 5 (83% in the 1978–88 cohort vs 96% in the post-1998 cohort) (figure 3A and figures E9–E10 in online appendix). There is a large change in the point estimate for the rate of change of lung function in the post-1998 (0.24%) compared with the 1988–98 cohort (–1% per year), such that the post-1998 cohort appears to be improving over the period of measurement. The three cohorts spanning the years 1948–1978 have a similar overall rate of decline around –0.3% per year, with an intercept at age 5 of 66%. Pancreatic insufficiency is associated with a significantly steeper rate of decline of lung function (–0.92% per year, 95% CI –1.7 to –0.3), as is acquisition of *Pseudomonas*

**Table 2** Estimates from final multivariate model

	Point estimate	Lower 95% CI	Upper 95% CI	p Value
Intercept at age 5 years	66.02	61.13	70.92	<0.001
CFRD	–2.47	–3.58	–1.37	<0.001
Age	–0.26	–0.49	–0.03	0.025
Cohort≥1948 (reference 1968)	1.20	–25.50	27.90	0.930
Cohort≥1958	–0.75	–10.01	8.51	0.874
Cohort≥1978	16.60	10.15	23.05	<0.001
Cohort≥1988	25.19	19.11	31.27	<0.001
Cohort≥1998	29.81	22.85	36.78	<0.001
Pancreatic sufficiency	2.78	–10.43	15.99	0.679
<i>Pseudomonas aeruginosa</i> infection	–0.51	–0.72	–0.29	<0.001
Age×cohort≥1948	–0.03	–0.67	0.61	0.920
Age×cohort≥1958	0.06	–0.23	0.34	0.699
Age×cohort≥1978	–0.72	–1.00	–0.44	<0.001
Age×cohort≥1988	–0.72	–1.09	–0.35	<0.001
Age×cohort≥1998	0.50	–0.41	1.42	0.280
Age×pancreatic sufficiency	0.98	0.29	1.67	0.005

CFRD, cystic fibrosis related diabetes.

**Figure 3** Effect of covariates on forced expiratory volume in 1 s as a percentage of predicted (%FEV<sub>1</sub>). (A) Birth cohort effect in the final model. There is clear separation between the three most recent birth cohorts, with a successive increase in the intercept term at age 5 years. (B) Effect of pancreatic insufficiency and *Pseudomonas* infection on the predicted population trajectory for a person born in the 1988–1998 cohort. CF, cystic fibrosis; PA, *Pseudomonas aeruginosa*.



infection ( $-0.5\%$  per year, 95% CI  $-0.72$  to  $-0.3$ ) (figure 3B and figure E8 in online appendix). CFRD is associated with a drop in intercept of  $-2.5\%$  (95% CI  $-3.6\%$  to  $-1.37\%$ ), but has no effect on the rate of decline of lung function.

## DISCUSSION

We describe a novel longitudinal modelling technique specifically aimed at analysing long sequences of repeated measurements, and apply this to %FEV<sub>1</sub> from a CF population. We show how this approach could be used to inform patient management, by aiding the interpretation of sudden changes in lung function, and by quantifying the predictive value of a baseline %FEV<sub>1</sub> measure up to 15 years later. At the population level, we show how our model can be used to quantify the effect of covariates on populations or sub-populations. Translation of these methods into clinical practice is important because people with CF are living longer, and we have shown how commonly applied approaches are unhelpful over long follow-up periods.

This study quantifies the short-term variability in %FEV<sub>1</sub> in this population (SD 6.3%), and demonstrates that %FEV<sub>1</sub> measures within individuals are correlated over time lags of 15 years or more. We have also explored the effect of previously studied risk factors for lung function decline in the Danish CF population, and have demonstrated significant effects of birth cohort, pancreatic status and *Pseudomonas* infection status.

The findings from this study have a number of clinical applications. Quantifying the variability in lung function measures is essential to make correct clinical interpretation.<sup>10</sup> Exploiting the unusually high frequency of data collection in Denmark, this study implies that on average a change in %FEV<sub>1</sub> of  $>13\%$  (ie, twice the error SD, to give a 95% confidence range) is likely to represent true within-patient variation over time (disease progression), whereas anything less than this could be due to short-term fluctuation, which may recover. Stanbrook *et al*<sup>21</sup> found a pooled within-subject %FEV<sub>1</sub> SD of 4.5% when measured over a 9-day period in 21 stable adults with CF. This population is different to the population in our study, who were measured regardless of clinical status, and one would therefore expect greater variability. Other studies have shown that people with CF, asthma and COPD have more short-term variability in lung function tests<sup>22–24</sup> and that more impaired lung function is associated with greater variability.<sup>25</sup>

Our model can be used to generate an underlying representation of an individual's 'true' lung function trajectory (figure 1B,D) that smoothes out the noise inherent in %FEV<sub>1</sub> measures. These smoothed traces could be used to inform clinical decision-

making—the model fit curves in figure 1 provide more realistic estimates of underlying lung function, and more valid criteria for clinical decisions. We propose that this model could be used to develop a real-time smoothing tool embedded in electronic patient records to aid clinical interpretation of spirometry data. We suggest that access to this information would provide some re-assurance to patients experiencing lower than expected lung function values, since lung function can recover quite dramatically, and these data suggest that a linear or stepwise decline in lung function over time is not the norm.

We have generated, for the first time to our knowledge, the variogram function for %FEV<sub>1</sub> in people with CF over long follow-up periods. This precisely quantifies how %FEV<sub>1</sub> measures are correlated over time. Furthermore we have done this for the whole CF population of Denmark. This quantifies the degree to which a baseline %FEV<sub>1</sub> measure can be used to predict subsequent %FEV<sub>1</sub> measures over long follow-up periods, and is likely to be of interest to clinicians and patients. We demonstrate a long-term correlation between levels of %FEV<sub>1</sub> within an individual. This suggests that there is long-term predictive value in a high %FEV<sub>1</sub> measure—people with CF with a high %FEV<sub>1</sub> at baseline are more likely to have a high %FEV<sub>1</sub> up to 15 years later than individuals with a lower baseline %FEV<sub>1</sub> (figure 2B). However, the predictive value of a %FEV<sub>1</sub> measure drops away rapidly over this period. We can say that on average a %FEV<sub>1</sub> reading today explains about 63% of the variability in %FEV<sub>1</sub> at 1 year, 40% at 3 years, and about 30% at 5 years.

This corroborates Rosenthal's study,<sup>26</sup> which found that baseline %FEV<sub>1</sub> explains 66% of the variability in %FEV<sub>1</sub> at 1 year, and Mastella *et al*'s study of European registry data in which differences in lung function at enrolment at age 5, categorised as mild, moderate or severe, tracked through the study to age 40.<sup>27</sup> Konstan *et al* also describe how a lower %FEV<sub>1</sub> for a given age can be used to characterise the aggressiveness of lung disease.<sup>28</sup> Other studies have shown a high %FEV<sub>1</sub> to be an independent risk factor for a greater rate of decline of %FEV<sub>1</sub> over the next few years.<sup>4 29</sup> This is not at odds with our findings here; a high %FEV<sub>1</sub> can be a risk factor for greater decline in the short term, while still being associated with a relatively higher %FEV<sub>1</sub> over the longer term.<sup>28</sup>

At the population level we show how our approach can be applied to quantify the effect of covariates on changes in lung function. Furthermore, the partitioning of the variability in %FEV<sub>1</sub> and the precise description of the correlation structure captured in the model provide important information for sample

size calculations in longitudinal clinical studies with %FEV<sub>1</sub> as an outcome. Increasingly longitudinal outcomes are being used in randomised control trials, and to undertake an a priori sample size calculation it is essential to have information on the correlation structure. Furthermore, our modelled %FEV<sub>1</sub> trace could be used as an outcome in its own right.

As with other studies of patients with CF,<sup>30</sup> there is a striking cohort effect evident in this population. The treatment of CF lung disease has been transformed over the period captured in this analysis, from 1969 to the present day. Particularly impressive is the improvement in lung function in the post-1998 cohort by comparison with preceding birth cohorts. Although patients in this group are early in their disease progression, the overall picture suggests that new therapeutic strategies are continuing to provide improvements in respiratory function in CF.

Our approach to modelling changes in %FEV<sub>1</sub> can be applied over long follow-up periods. This is in contrast to the widely used random intercept and slope approach that has been applied in studies of CF and COPD over short-term<sup>4 27 31 32</sup> and longer-term follow-up periods.<sup>10 11 14 33</sup> The development and testing of the new approach is facilitated by the nature of the Danish CF register—to our knowledge there are no other datasets that contain such frequent (monthly) measures of lung function on individuals measured over very long periods (up to 31.5 years). However, the fact that the data are from Denmark does not influence the validity of the methods we have described, since these are essentially context free. Furthermore, this method does not exploit any features of our data that are unique to CF, and is equally applicable to other clinical areas that generate long sequences of repeated measurements. As a next step we recommend that this method be applied to longitudinal data collected in other CF registries, such as the UK, to clarify how robust this approach is in terms of predicting changes in %FEV<sub>1</sub> over time, and to better understand how this might inform clinical decision making. Future research could explore the utility of our proposed model in other diseases such as COPD.

A limitation of this study is the likely influence of survivor bias on lung function estimates in the earlier birth cohorts. In the 1948–1978 period, the intercept at age 5 appears significantly lower than in the other cohorts, but there is also a shallower rate of decline of lung function. This is likely to be due to the incomplete capture of patients in earlier cohorts, with censoring due to death leaving only the more stable survivors. This is a common problem in datasets of this type.<sup>34</sup> Fitting the model by maximum likelihood automatically corrects for selection bias that depends on a patient's observed lung function measurements prior to death, although not for any additional dependence on unmeasured features of their lung function trajectory.<sup>15 19</sup>

Pancreatic sufficiency had an important effect on the overall rate of decline of lung function (+0.9% per year). In Konstan's study<sup>4</sup> pancreatic sufficiency was the most important protective factor in the age group 6–8 years (+1.33% per year). The small number of pancreatic-sufficient individuals in the Danish dataset (n=20, 5%) have a notably different lung function phenotype, maintaining near-normal lung function over the period of follow-up (see plot in online appendix). The onset of *Pseudomonas* infection was associated with a significant increase in the rate of decline of lung function, by around –0.5% per year, similar to that reported in the study by Konstan, in which *Pseudomonas* colonisation was associated with an increased rate of decline of FEV<sub>1</sub> of –0.31% per year in the 6–8-year-old age group, and –0.22 in the 9–12-year-old age group.<sup>4</sup>

In conclusion, our modelling approach provides a more realistic estimate of the %FEV<sub>1</sub> trajectory in CF, which could be applied in real time to help clinicians interpret the significance of changes in %FEV<sub>1</sub>. Furthermore, our approach quantifies the predictive value of a baseline %FEV<sub>1</sub> measure, over three decades. This method is equally applicable to the longitudinal assessment of %FEV<sub>1</sub> in other lung diseases, and can enable more robust comparisons of populations, including groups studied in clinical trials. As people are now living for many decades with these diseases, the development of tools to better understand the natural history of this important outcome will be essential for improved clinical care, as well as being a key research priority.<sup>1</sup>

**Acknowledgements** We thank Professor Peter Oluf Schiøtz for his support in accessing the data for this analysis.

**Contributors** DTR, MMW, FD, TP, RLS and PD conceived and designed the study. TP and HVO collected the data. DTR undertook the analysis and PD supervised analysis. DTR, MMW, RLS and PD interpreted the results and drafted the paper. All authors contributed to and approved the final draft for publication.

**Funding** This work was supported by an MRC Population Health Scientist Fellowship to DTR (G0802448). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests** None.

**Ethics approval** The study was approved by the Danish Data inspectorate (Datatilsynet). Danish CF registry data were used, analysed anonymously.

**Provenance and peer review** Not commissioned; externally peer reviewed.

## REFERENCES

- Holgate ST. Priorities for respiratory research in the UK. *Thorax* 2007;**62**:5–7.
- Davies JC, Alton EW. Monitoring respiratory disease severity in cystic fibrosis. *Respir Care* 2009;**54**:606–17.
- Rabe KF, Hurd S, Anzueto A, et al. Global Initiative for Chronic Obstructive Lung Disease. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med* 2007;**176**:532–55.
- Konstan MW, Morgan WJ, Butler SM, et al. Risk factors for rate of decline in forced expiratory volume in one second in children and adolescents with cystic fibrosis. *J Pediatr* 2007;**151**:134–9, 139.e1.
- Konstan MW, Schluchter MD, Xue W, et al. Clinical use of ibuprofen is associated with slower FEV1 decline in children with cystic fibrosis. *Am J Respir Crit Care Med* 2007;**176**:1084–9.
- Ramsey BW, Dorkin HL, Eisenberg JD, et al. Efficacy of aerosolized tobramycin in patients with cystic fibrosis. *N Engl J Med* 1993;**328**:1740–6.
- Kerem E, Reisman J, Corey M, et al. Prediction of mortality in patients with cystic fibrosis. *N Engl J Med* 1992;**326**:1187–91.
- George PM, Banya W, Pareek N, et al. Improved survival at low lung function in cystic fibrosis: cohort study from 1990 to 2007. *BMJ* 2011;**342**:d1008.
- Corey M. Power considerations for studies of lung function in cystic fibrosis. *Proc Am Thorac Soc* 2007;**4**:334–7.
- Hnizdo E, Yu L, Freyder L, et al. The precision of longitudinal lung function measurements: monitoring and interpretation. *Occup Environ Med* 2005;**62**:695–701.
- Stern DA, Morgan WJ, Wright AL, et al. Poor airway function in early infancy and lung function by age 22 years: a non-selective longitudinal cohort study. *Lancet* 2007;**370**:758–64.
- van Diemen C, Postma D, Siedlinski M, et al. Genetic variation in TIMP1 but not MMPs predict excess FEV1 decline in two general population-based cohorts. *Respir Res* 2011;**12**:57.
- Harber P, Tashkin DP, Simmons M, et al. Effect of occupational exposures on decline of lung function in early chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* 2007;**176**:994–1000.
- Corey M, Edwards L, Levison H, et al. Longitudinal analysis of pulmonary function decline in patients with cystic fibrosis. *J Pediatr* 1997;**131**:809–14.
- Diggle P, Heagerty P, Liang KY, et al. *Analysis of Longitudinal Data*. 2nd edn. Oxford: Oxford University Press, 2002.
- Miller MR, Hankinson J, Brusasco V, et al. Standardisation of spirometry. *Eur Respir J* 2005;**26**:319–38.
- Wang X, Dockery DW, Wypij D, et al. Pulmonary function between 6 and 18 years of age. *Pediatr Pulmonol* 1993;**15**:75–88.
- Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. *Am J Respir Crit Care Med* 1999;**159**:179–87.
- Fitzmaurice GM, Laird NM, Ware JH. *Applied Longitudinal Analysis*. Hoboken, New Jersey: Wiley, 2004.

20. **Laird NM**, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982;**38**:963–74.
21. **Stanbrook MB**, Corey M, Tullis DE. The repeatability of forced expiratory volume measurements in adults with cystic fibrosis. *Chest* 2004;**125**:150–5.
22. **Hruby J**, Butler J. Variability of routine pulmonary function tests. *Thorax* 1975;**30**:548–53.
23. **Pennock BE**, Rogers RM, McCaffree DR. Changes in measured spirometric indices. What is significant? *Chest* 1981;**80**:97–9.
24. **Cooper PJ**, Robertson CF, Hudson IL, *et al*. Variability of pulmonary function tests in cystic fibrosis. *Pediatr Pulmonol* 1990;**8**:16–22.
25. **Enright PL**, Beck KC, Sherrill DL. Repeatability of spirometry in 18,000 adult patients. *Am J Respir Crit Care Med* 2004;**169**:235–8.
26. **Rosenthal M**. Annual assessment spirometry, plethysmography, and gas transfer in cystic fibrosis: do they predict death or transplantation. *Pediatr Pulmonol* 2008;**43**:945–52.
27. **Mastella G**, Rainisio M, Harms HK, *et al*. Allergic bronchopulmonary aspergillosis in cystic fibrosis. A European epidemiological study. Epidemiologic Registry of Cystic Fibrosis. *Eur Respir J* 2000;**16**:464–71.
28. **Konstan MW**, Wagener JS, VanDevanter DR. Characterizing aggressiveness and predicting future progression of CF lung disease. *J Cyst Fibros* 2009;**8**(Suppl 1): S15–19.
29. **Vandevanter DR**, Wagener JS, Pasta DJ, *et al*. Pulmonary outcome prediction (POP) tools for cystic fibrosis patients. *Pediatr Pulmonol* 2010;**45**:1156–66.
30. **Elborn JS**, Shale DJ, Britton JR. Cystic fibrosis: current survival and population estimates to the year 2000. *Thorax* 1991;**46**:881–5.
31. **Vestbo J**, Sorensen T, Lange P, *et al*. Long-term effect of inhaled budesonide in mild and moderate chronic obstructive pulmonary disease: a randomised controlled trial. *Lancet* 1999;**353**:1819–23.
32. **MacLean JE**, Atenafu E, Kirby-Allen M, *et al*. Longitudinal decline in lung volume in a population of children with sickle cell disease. *Am J Respir Crit Care Med* 2008;**178**:1055–9.
33. **Kohansal R**, Martinez-Camblor P, Agusti A, *et al*. The natural history of chronic airflow obstruction revisited: an analysis of the Framingham offspring cohort. *Am J Respir Crit Care Med* 2009;**180**:3–10.
34. **Frederiksen B**, Lannig S, Koch C, *et al*. Improved survival in the Danish center-treated cystic fibrosis patients: results of aggressive treatment. *Pediatr Pulmonol* 1996;**21**:153–8.

### Author's response: understanding the natural progression in %FEV<sub>1</sub> decline in patients with cystic fibrosis: a longitudinal study

We thank Professor Miller for his comments<sup>1</sup> regarding our paper,<sup>2</sup> in which we outline a novel approach to modelling repeated lung function measures in people with cystic fibrosis (CF) over long follow-up periods. We agree that it would be interesting to apply our methodology using alternative methods of expressing lung function across the age range, and reiterate that our approach can be usefully applied to any clinical outcome measured repeatedly over extended periods.

While the merits of different approaches are appreciated, we chose to model forced expiratory volume in 1 s as a percentage of predicted (%FEV<sub>1</sub>) for a number of reasons. First, %FEV<sub>1</sub> is still currently recognised as a key outcome measure in CF<sup>3</sup> as it is predictive of survival, and is currently an important criterion in international lung transplant guidelines.<sup>4</sup> Second, %FEV<sub>1</sub> has been modelled previously over long follow-up periods, across the paediatric and adult age range in CF using a random-intercept and slope approach,<sup>5</sup> and we wanted to compare our method with this. Third, standardised %FEV<sub>1</sub> was the most commonly collected outcome measure in the Danish CF registry.

While we agree that use of recently developed all-age equations, such as those published by Stanojevic<sup>6</sup> or Quanjer<sup>7</sup> would be advantageous in avoiding arbitrary breaks, for the purposes of this analysis, we adhered to the approach that is currently used in Denmark to facilitate comparisons with previously published data. Irrespective of the precise outcome

used, our finding that the error in repeated measurements of %FEV<sub>1</sub> within individuals is large (average within-person SD of 6 percentage points) remains valid. Furthermore, our approach provides more realistic estimates of the underlying lung-function trajectory of people with chronic lung disease, by acknowledging the imprecision in individual measurements over time, and the correlation structure of repeated measurements on the same individual, issues that have all too often been disregarded in the past.

David Taylor-Robinson,<sup>1</sup> Margaret Whitehead,<sup>1</sup>  
Finn Diderichsen,<sup>2</sup> Hanne Veberth Olesen,<sup>3</sup>  
Tania Pressler,<sup>4</sup> Rosalind Smyth,<sup>5</sup> Peter Diggle<sup>6</sup>

<sup>1</sup>Department of Public Health and Policy, University of Liverpool, Liverpool, UK

<sup>2</sup>Department of Social Medicine, University of Copenhagen, Copenhagen, Denmark

<sup>3</sup>Cystic Fibrosis Center, Aarhus University Hospital, Aarhus, Denmark

<sup>4</sup>Cystic Fibrosis Center, Rigshospitalet, Copenhagen, Denmark

<sup>5</sup>Institute of Child Health, UCL, London, UK

<sup>6</sup>Institute of Infection and Global Health, University of Liverpool, Liverpool, UK

**Correspondence to** Dr David Taylor-Robinson, Department of Public Health and Policy, University of Liverpool, Liverpool, UK; [dctr@liv.ac.uk](mailto:dctr@liv.ac.uk)

**Funding** DTR is supported by an MRC Population Health Scientist Fellowship (G0802448).

**Competing interests** None.

**Provenance and peer review** Not commissioned; internally peer reviewed.

**To cite** Taylor-Robinson D, Whitehead M, Diderichsen F, *et al.* *Thorax* 2013;**68**:294–295.

Published Online First 4 December 2012



► <http://dx.doi.org/10.1136/thoraxjnl-2012-202838>

*Thorax* 2013;**68**:294–295.  
doi:10.1136/thoraxjnl-2012-202954

#### REFERENCES

- 1 Miller M. FEV<sub>1</sub> decline in cystic fibrosis. *Thorax* 2012 doi:10.1136/thoraxjnl-2012-202838.
- 2 Taylor-Robinson D, Whitehead M, Diderichsen F, *et al.* Understanding the natural progression in %FEV<sub>1</sub> decline in patients with cystic fibrosis: a longitudinal study. *Thorax* 2012;**67**:860–6.
- 3 Rosenfeld M, Coates A, Corey M, *et al.* Task force to evaluate choice of spirometric reference equations for the National Patient Registry: summary and recommendations. Cystic Fibrosis Foundation Registry Committee, 1 Oct 2005.
- 4 Orens JB, Estenne M, Arcasoy S, *et al.* International guidelines for the selection of lung transplant candidates: 2006 update—a consensus report from the Pulmonary Scientific Council of the International Society for Heart and Lung Transplantation. *J Heart Lung Transplant* 2006;**25**:745–55.
- 5 Corey M, Edwards L, Levison H, *et al.* Longitudinal analysis of pulmonary function decline in patients with cystic fibrosis. *J Pediatrics* 1997;**131**:809–14.

- 6 Stanojevic S, Wade A, Stocks J, *et al.* Reference ranges for spirometry across all ages: a new approach. *Am J Respir Crit Care Med* 2008;177:253–60.
- 7 Quanjer PH, Stanojevic S, Cole TJ, *et al.* Multi-ethnic reference values for spirometry for the 3–95 year age range: the global lung function 2012 equations. *Eur Respir J* Published Online First: 27 Jun 2012. doi: 10.1183/09031936.00080312.

## EDITORIALS

## Health inequalities and cystic fibrosis

Even for genetic diseases, social conditions are still important determinants of outcome

David Taylor-Robinson *Medical Research Council population health scientist*<sup>1</sup>, Michael S Schechter *associate professor*<sup>2</sup>

<sup>1</sup>Department of Public Health and Policy, University of Liverpool, Liverpool L69 3GB, UK; <sup>2</sup>Department of Pediatrics, Emory University School of Medicine and Children's Healthcare of Atlanta, Atlanta, GA, USA

Cystic fibrosis is the most common serious inherited disease in white populations. Intensive support from family and healthcare services is needed from the time of diagnosis onwards, and most patients die prematurely from respiratory failure. Survival has dramatically improved over successive birth cohorts, such that British children born in the 21st century will have a median survival of over 50 years.<sup>1</sup> However, disease progression and survival still vary greatly, mostly as a result of social and environmental, rather than genetic, determinants.<sup>2</sup> It has been known for more than 20 years that people with cystic fibrosis from socioeconomically disadvantaged backgrounds die younger than those in more advantaged positions.<sup>3</sup>

In a linked observational study (doi:10.1136/bmj.d4662), Barr and colleagues used death registration data in England and Wales to show that this socioeconomic divide in premature mortality in cystic fibrosis has persisted with no substantial narrowing for over four decades.<sup>4</sup> They also show that female patients continue to die at a younger age than men. What can we learn from this, and what are the implications for policy and for clinicians?

Cystic fibrosis offers a valuable case for understanding how health inequalities develop. It is an autosomal recessive disease with an asymptomatic (and, until recently, undetectable) carrier state, so unlike many other diseases, socioeconomic status does not influence who gets the disease. Inequalities related to socioeconomic status result from the different patterns of exposure to harmful and protective or therapeutic influences that occur over the course of people's lives. Studies from the United States and United Kingdom show that significant inequalities in key intermediate outcomes in cystic fibrosis, such as growth and lung function, begin early in childhood and persist over time.<sup>5,6</sup> The early appearance and persistence of inequalities support the need for interventions that are targeted at the early (and perhaps prenatal) years and reinforce the importance of screening for cystic fibrosis in newborns. This, incidentally, is also true for sex related inequalities, which are probably caused as much by socially determined gender roles as by biologically determined sex characteristics.<sup>7</sup>

A key question for practising clinicians is what role healthcare delivery plays in mitigating or potentiating health inequalities in cystic fibrosis. In the US, studies have failed to identify important socioeconomic status related differences in the use of chronic treatment for cystic fibrosis, treatment of pulmonary exacerbations, or hospital admissions,<sup>2,5,8</sup> although there is some evidence that affluent groups may be earlier recipients of newly developed cystic fibrosis drugs.<sup>8</sup> The adoption of system based methods to optimise consistency in the use of best care practices might help to minimise variations in prescribed care.<sup>9</sup> Other than tackling any residual socioeconomic status related differences in access and provision of healthcare to patients with cystic fibrosis, how else can inequalities be reduced?

Some underused tools are available to cystic fibrosis clinicians. One obvious target for action is to protect newly diagnosed children from environmental tobacco smoke. There are striking and persistent differences in the prevalence of smoking according to socioeconomic status.<sup>10</sup> Exposure to environmental tobacco smoke is associated with poorer growth and lung function in cystic fibrosis,<sup>2</sup> and it may be the most important explanatory factor for inequalities related to socioeconomic status in this disease.<sup>11</sup> Early identification of family members who smoke, and appropriate counselling and referral to smoking cessation services, would be an effective intervention for all patients, regardless of social position. This should be coupled with support to develop disease self management skills in the patient and the family,<sup>12</sup> in addition to targeted input from social workers who work with the multidisciplinary cystic fibrosis team.

Ultimately, however, although individually focused interventions may have some limited success, the more effective long term solution to health inequalities in people with cystic fibrosis and in the general population is likely to be one that takes broader action to tackle the social determinants of health. These are the "conditions in which we are born, grow up, work, and live," and they include income and income distribution, education, employment and working conditions, housing, food insecurity, race and ethnicity, and sex and gender roles. These factors provide a particularly important context for a family dealing

with the stresses of caring for a child with a complex chronic illness like cystic fibrosis over a lifetime.

The evidence is clear, unfortunately, that we have made little progress over the past few decades in reducing health inequalities generated by social gradients,<sup>11</sup> and this evidence is strengthened by Barr and colleagues' study. Future research should assess the mechanisms that generate socioeconomic status related inequalities and the interventions that are most likely to reduce them, bearing in mind that investigating interventions at the population level is likely to have greater impact. The current political discourse in the US suggests that any insight into effective social interventions is unlikely to come from that side of the Atlantic; the discussion on the UK side, inspired by documents such as the Marmot report,<sup>11</sup> will hopefully provide a first step towards investigating and putting into practice the most effective ways to reduce the socioeconomic gradient in health.

Competing interests: All authors have completed the ICMJE uniform disclosure form at [www.icmje.org/coi\\_disclosure.pdf](http://www.icmje.org/coi_disclosure.pdf) (available on request from the corresponding author) and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Provenance and peer review: Commissioned; not externally peer reviewed.

- 1 Dodge JA, Lewis PA, Stanton M, Wilsher J. Cystic fibrosis mortality and survival in the UK: 1947-2003. *Eur Respir J* 2007;29:522.
- 2 Schechter MS. Non-genetic influences on cystic fibrosis lung disease: the role of sociodemographic characteristics, environmental exposures, and healthcare interventions. *Semin Resp Crit Care Med* 2003;24:639-52.
- 3 Britton JR. Effects of social class, sex, and region of residence on age at death from cystic fibrosis. *BMJ* 1989;298:483-7.
- 4 Barr HL, Britton JR, Smyth AR, Fogarty AW. Association between socioeconomic status, sex, and age at death from cystic fibrosis in England and Wales (1959 to 2008): cross sectional study. *BMJ* 2011;343:d4662.
- 5 Schechter MS, Shelton BJ, Margolis PA, Fitzsimmons SC. The association of socioeconomic status with outcomes in cystic fibrosis patients in the United States. *Am J Respir Crit Care Med* 2001;163:1331-7.
- 6 Taylor-Robinson D, Diggle P, Whitehead M, Smyth R. The effect of social deprivation on weight in the UK cystic fibrosis population. *J Cystic Fibrosis* 2011;10(suppl 1):S72.
- 7 Willis E, Miller R, Wyn J. Gendered embodiment and survival for young people with cystic fibrosis. *Soc Sci Med* 2001;53:1163-74.
- 8 Schechter MS, McColley SA, Silva S, Haselkorn T, Konstan MW, Wagener JS. Association of socioeconomic status with the use of chronic therapies and healthcare utilization in children with cystic fibrosis. *J Pediatr* 2009;155:634-9e1-4.
- 9 Schechter MS, Margolis P. Improving subspecialty healthcare: lessons from cystic fibrosis. *J Pediatr* 2005;147:295-301.
- 10 Marmot M. Fair society, healthy lives. 2010. [www.marmotreview.org/AssetLibrary/pdfs/Reports/FairSocietyHealthyLivesExecSummary.pdf](http://www.marmotreview.org/AssetLibrary/pdfs/Reports/FairSocietyHealthyLivesExecSummary.pdf).
- 11 Collaco JM, Vanscoy L, Bremer L, McDougal K, Blackman SM, Bowers A, et al. Interactions between secondhand smoke and genes that affect cystic fibrosis lung disease. *JAMA* 2008;299:417-24.
- 12 Goldman DP, Smith JP. Can patient self-management help explain the SES health gradient? *Proc Natl Acad Sci USA* 2002;99:10929-34.

Cite this as: *BMJ* 2011;343:d4818

© BMJ Publishing Group Ltd 2011

### S18.3

## THE EFFECT OF SOCIOECONOMIC STATUS ON OUTCOMES IN CF

David Taylor-Robinson, MB Ch.B.(Hons), B.Sc.(Hons), MPH<sup>1</sup>, Hanne V. Olesen<sup>2</sup>, Tania Pressler<sup>3</sup>, Karsten Thielen<sup>4</sup>, Finn Diderichsen<sup>4</sup>, Peter Diggle<sup>1,5</sup>, Rosalind Smyth<sup>1</sup> and Margaret Whitehead<sup>1</sup>

1. University of Liverpool, Liverpool, United Kingdom; 2. Aarhus University Hospital, Aarhus, Denmark; 3. Rigshospitalet, Copenhagen, Denmark; 4. University of Copenhagen, Copenhagen, Denmark; 5. Lancaster University, Lancaster, Denmark

**Overview:** This study investigates the effect of socioeconomic status (SES) on outcomes in people with cystic fibrosis (CF) in the UK and Denmark. People with CF from socio-economically disadvantaged backgrounds die younger than those in more advantaged social positions in the UK and the U.S. The key challenge is to understand how and when these inequalities develop, in order to identify promising options for intervention. We demonstrate the importance of social factors for key outcomes in CF, in the context of the universal health systems in the UK and Denmark, and contrast these findings with those from the US and other settings.

**Background:** Studies across the world have consistently shown that people from socio-economically disadvantaged backgrounds experience worse health than those in more socio-economically advantaged positions [1]. In the UK and internationally, policies have been implemented to try to reduce these inequalities, with limited success [2]. In order to develop more effective interventions we need a better understanding of how these health differences are generated and maintained. CF is a valuable case for understanding how and when health inequalities develop, since unlike other childhood respiratory disorders CF does not discriminate by social class: SES does not affect disease risk but does affect the health outcomes of having CF [3].

The improvement in survival over successive birth cohorts in CF has been striking, but these improvements have not been shared evenly, both within and between countries [4]. Evidence from the U.S. indicates higher survival rates in the 1980s and 1990s among more advantaged socioeconomic groups, measured by Medicaid status and area-based income, compared with their less advantaged counterparts [5-7]. For instance the adjusted risk of death was almost four times higher in CF patients with Medicaid cover, a surrogate for low socioeconomic status, compared to those without Medicaid cover [5]. In the UK a cross-sectional study by Britton found a consistent trend from 1959 to 1986 toward higher age at death in CF patients in non-manual, compared with manual occupations [8]. Barr and colleagues recently updated this study, using death registration data in England and Wales, and showed that the socioeconomic divide in premature mortality in CF has persisted with no substantial narrowing for over four decades [9].

There are significant differences in intermediate CF outcomes such as %FEV1 and weight centile in young people with CF in the U.S. For instance Schechter et al. found a cross sectional difference of 6.7% in %FEV1 by Medicaid status, which increased to 9.2% with adjustment for various confounders [5]. Furthermore O'Connor et al report a stepped social gradient in %FEV1 and weight centile, with an absolute difference of 5.5% between highest and lowest income quintiles in the under 18 age group [7]. A key issue is the role of the healthcare system in mitigating or perpetuating inequalities in CF. Michael Schechter's studies from the U.S. have shown mixed patterns of access to treatment by SES [10,11]. Young people on Medicaid appear more likely to access care (sick hospital visits, chronic therapies and IV antibiotics). Children living in low-income areas are more likely to receive oral nutritional supplementation, but less likely to receive macrolide therapy, and IV antibiotics. The authors conclude that inequalities in CF outcomes are not explained by differences in healthcare or chronic therapy use. In their study of the effect of socioeconomic status on hospitalisation rates in CF in Ontario Stephenson et al suggest that the provision of universal healthcare in Canada may explain the lack of a socioeconomic differential in hospitalisations [12].

To investigate these issues in the UK and Denmark, we undertook retrospective cohort studies of population level cohorts.

**Methods:** For the UK analysis we studied around 9000 people with CF aged under 40 years, with data captured at 58,000 annual reviews between 1996 and 2010. Standardised census-based indices of multiple deprivation (IMD) from the UK constituent countries were used as small area measures of SES. In Denmark the analysis captured 479 patients seen between 1969 and 2010 at the two CF centers in Denmark, with data measured on a monthly basis on around 70,000 occasions. The Danish dataset was linked to population level registers in order to collect individual level socioeconomic data, on patients and their parents (e.g., employment, income, healthcare expenditure). In both countries we explored the effects of socioeconomic status measures on key clinical outcomes (%FEV1, growth and nutritional measures), healthcare use (access to care, use of chronic therapies), and social outcomes (employment and educational outcomes). The analyses employed longitudinal data modelling techniques to examine changes in outcomes over time in groups and

individuals whilst allowing for correlation within patients, and potentially informative missing values [13,14].

**Findings:** We have, for the first time, characterised key outcomes over time in people with CF in the UK and Denmark, and explored the impact of social deprivation, as well as other covariates. Our findings, to date, have shown important differences in outcomes by socioeconomic status in the UK, which appear early, and vary across the life-course, perhaps as a consequence of good access to treatment in children from families living in areas with greater social deprivation. The early appearance of inequalities supports the need for interventions that are targeted at the early years and reinforce the importance of screening for cystic fibrosis in newborns. Research to identify interventions to address the early appearance of inequalities needs to be focussed on the early years.

**References:** 1. CSDH. Closing the gap in a generation: health equity through action on the social determinants of health. (2008)[http://whqlibdoc.who.int/publications/2008/9789241563703\\_eng.pdf](http://whqlibdoc.who.int/publications/2008/9789241563703_eng.pdf).

2. Mackenbach JP. Can we reduce health inequalities? An analysis of the English strategy (1997-2010). *J Epidemiol Community Health* 2011;65:568-575.

3. Taylor-Robinson D, Schechter MS. Health inequalities and CF. *BMJ*. 2011;343:d4818.

4. Fogarty A, Hubbard R, Britton J. International comparison of median age at death from CF. *Chest*. 2000;117:1656-1660.

5. Schechter MS, Shelton BJ, Margolis PA, Fitzsimmons SC. The association of socioeconomic status with outcomes in CF patients in the U.S. *Am J Respir Crit Care Med*. 2001;163:1331-1337.

6. Schechter MS. Non-genetic influences on CF lung disease: the role of sociodemographic characteristics, environmental exposures and health care interventions. *Pediatr Pulmonol*. 2004;Suppl 26:82-85.

7. O'Connor GT, Quinton HB, Kneeland T, Kahn R, Lever T, et al. Median household income and mortality rate in CF. *Pediatrics*. 2003;111:e333-339.

8. Britton JR. Effects of social class, sex, and region of residence on age at death from CF. *BMJ*. 1989;298:483-487.

9. Barr HL, Britton J, Smyth AR, Fogarty AW. Association between socioeconomic status, sex, and age at death from CF in England and Wales (1959 to 2008): cross sectional study. *BMJ*. 2011 Aug 23;343:d4662.

10. Schechter MS, McColley SA, Silva S, Haselkorn T, Konstan MW, et al. Association of socioeconomic status with the use of chronic therapies and healthcare utilization in children with CF. *J Pediatr*. 2009;155:634-639.e1-4.

11. Schechter MS, McColley SA, Regelman W, Millar SJ, Pasta DJ, et al. Socioeconomic status and the likelihood of antibiotic treatment for signs and symptoms of pulmonary exacerbation in children with CF. *J Pediatr*. 2011;159:819-824 e811.

12. Stephenson A, Hux J, Tullis E, Austin PC, Corey M, et al. Socioeconomic status and risk of hospitalization among individuals with CF in Ontario, Canada. *Pediatr Pulmonol*. 2011 Apr;46(4):376-84.

13. Diggle P, Heagerty P, Liang K-Y, Zeger SL (2002) *Analysis of Longitudinal Data* (second edition). Oxford: Oxford University Press.

14. Taylor-Robinson D, Whitehead M, Diderichsen F, Olesen HV, Pressler T, et al. Understanding the natural progression in %FEV1 decline in patients with CF: a longitudinal study. *Thorax*. 2012 May 3. [Epub ahead of print]