

PROJECTIONS OF LUNG CANCER INCIDENCE:  
MODELS INCORPORATING SMOKING BEHAVIOUR  
AND THEIR APPLICATION TO

LIVERPOOL  
HEALTH SERVICE ISSUES  
UNIVERSITY  
LIBRARY



Juanita Hatcher

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

May 1996

## TABLE OF CONTENTS

ABSTRACT	i
ACKNOWLEDGEMENTS	ii
LIST OF FIGURES AND TABLES	iv
LIST OF APPENDICES	xiii
1. INTRODUCTION	1
1.1 Lung Cancer Incidence	3
1.2 Projections	7
1.3 Health of the Nation	10
1.4 Service Needs	12
1.5 Outline of Thesis	13
2. LITERATURE REVIEW	15
2.1 Uses of Projections	19
2.2 Methodology for Projections	21
2.2.1 Statistical Modelling	22
2.2.2 Time Series	27
2.2.2.1 Age-standardised Incidence Rates	30
2.2.2.2 Age-Specific Rates	32
2.2.2.3 Age-Period-Cohort Modelling	37
2.2.3 Inclusion of Risk Factor Information in the Models	47
2.3 Aetiology of Lung Cancer	57
2.3.1 Cigarette Smoking	58
2.3.2 Occupational Exposure	65
2.3.3 Air Pollution	68
2.3.4 Diet	69

2.3.5 Social Class	71
2.3.6 Summary of Aetiology of Lung Cancer	72
2.4 Effects of intervention	73
2.5 Summary	76
<b>3. INFORMATION REQUIREMENTS FOR MODELLING LUNG CANCER INCIDENCE RATES</b>	<b>79</b>
3.1 Geographic Areas	81
3.2 Mersey Regional Cancer Registry	85
3.3 Population Estimates	91
3.3.1 Population Estimates for 1951-1988	91
3.3.2 Populations Projections up to the Year 2011	93
3.4 Risk Factor Information	94
3.5 Service Needs Information	101
3.6 Summary	102
<b>4. TRENDS IN LUNG CANCER INCIDENCE IN MERSEY REGION AND SIMPLE METHODS OF PROJECTIONS</b>	<b>105</b>
4.1 Trends in Lung Cancer Incidence Frequencies	108
4.1.1 Methods	108
4.1.2 Results	108
4.2 Trends in Crude Lung Cancer Incidence Rates	109
4.2.1 Methods	109
4.2.2 Results	111
4.3 Calculation of Age-Specific Lung Cancer Incidence Rates	113
4.4 Trends in Age-Standardised Lung Cancer Incidence Rates	117
4.4.1 Methods	117

4.4.2 Results	118
4.5 Trends in Broad-Band Age-Specific Lung Cancer Incidence Rates	120
4.5.1 Methods	120
4.5.2 Results	121
4.6 Age-Period Plots	123
4.6.1 Methods	123
4.6.2 Results	124
4.7 Age-Cohort Plots	127
4.7.1 Methods	127
4.7.2 Results	128
4.8 Summary of Trend Analyses	130
4.9 Projections	131
4.10 Projections for Age-Standardised Lung Cancer Incidence Rates	132
4.10.1 Methods	132
4.10.2 Results	134
4.10.3 Health of the Nation Targets	137
4.10.4 Service Needs	137
4.11 Projections for Broad-Band Age-Specific Lung Cancer Incidence Rates	138
4.11.1 Methods	138
4.11.2 Results	141
4.11.3 Health of the Nation Targets	143
4.11.4 Service Needs	143
4.12 Summary of Projections	143



5. TRENDS IN SMOKING BEHAVIOUR	147
5.1 Methods	149
5.2 Results	152
5.2.1 Trends in Manufactured Cigarettes Smoked per Person per Year	152
5.2.2 Trends in the Percentage of People who Smoke Manufactured Cigarettes	156
5.2.3 Trends in the Number of Cigarettes Smoked per Smoker per Week	159
5.2.4 Trends in the Percentage of Smokers who Smoke Plain Cigarettes	161
5.2.5 Trends in Sales Adjusted Average Tar Content of Cigarettes	162
5.3 Summary	163
6. MODELS BASED ON AGE-STANDARDISED LUNG CANCER INCIDENCE AND INFORMATION ON SMOKING BEHAVIOUR	166
6.1 Methods for Fitting Models incorporating Information on Smoking Behaviour	169
6.2 Results of Fitting the Models	173
6.3 Projections	179
6.3.1 Methods	179
6.3.2 Results	181
6.3.3 Health of the Nation Targets	184
6.3.4 Service Needs	185
6.4 Summary	186

7. MODELS BASED ON BROAD-BAND AGE-SPECIFIC LUNG CANCER INCIDENCE RATES AND INFORMATION ON SMOKING BEHAVIOUR	188
7.1 Methods for Fitting Models incorporating Information on Smoking Behaviour	191
7.2 Results of Fitting the Models	195
7.3 Projections	202
7.3.1 Methods	202
7.3.2 Results	203
7.3.3 Health of the Nation Targets	206
7.3.4 Service Needs	207
7.4 Summary	208
8. AGE-PERIOD-COHORT MODELLING	210
8.1 Methods for Age-Period-Cohort Variation	213
8.1.1 Fitting the Models	226
8.2 Results	230
8.2.1 Male Lung Cancer	231
8.2.2 Female Lung Cancer	238
8.2.3 Summary of Model Fitting	244
8.3 Projections	245
8.3.1 Methods	245
8.3.2 Results	249
8.3.3 Health of the Nation Targets	255
8.3.4 Service Needs	255
8.3.5 Summary of Projections	257

9. MODELS BASED ON AGE-PERIOD-COHORT ANALYSIS	260
INCORPORATING INFORMATION ON SMOKING BEHAVIOUR	
9.1 Data Used	262
9.2 Inclusion of Average Tar Content in the Model	264
9.2.1 Methods	264
9.2.2 Results	265
9.3 Inclusion of Cohort Specific Smoking Information	267
9.3.1 Methods	267
9.3.2 Results	270
9.4 Projections	281
9.4.1 Determination of the Average Tar Content	282
9.4.2 Determination of the Information on Smoking Behaviour	283
9.4.3 Determination of the Age-specific Lung Cancer Incidence for 1991, 2001 and 2011	283
9.5 Results	286
9.5.1 Average Tar Content	286
9.5.2 Extrapolation of the Cohort Smoking Values	287
9.5.3 Health of the Nation Targets for Smoking	289
9.5.4 Projected Lung Cancer Incidence	290
9.6 Health of the Nation Targets	300
9.7 Service Needs	302
9.8 Summary of Projections	306
10. DISCUSSION	310
10.1 Methods of Projection	313
10.2 Comparison of Projections	325
10.3 Implications for Health of the Nation	331
10.4 Implications for Service Needs	338

10.5 Conclusions

343

REFERENCES

346

APPENDICES

372



**Projections of Lung Cancer Incidence: Models  
incorporating Smoking Behaviour and their Application to  
Health Service Issues      *Juanita Hatcher***

In Mersey Region lung cancer is the most commonly occurring cancer in males and incidence is decreasing over time. In females lung cancer is the third most common cancer and incidence is increasing. The Health of the Nation strategy set targets for the reduction of lung cancer mortality of 30% in males and 15% in females. To attain these targets, targets for the reduction in smoking have been set. Up to 90% of all lung cancers are due to smoking. Accurate projections of future cancer incidence would assist the evaluation of the Health of the Nation strategy and the planning of cancer services. This thesis presents projection models, based on past trends of lung cancer incidence, that incorporate information on smoking behaviour.

Data on incident lung cancer cases for 1951 to 1988 were obtained from the Mersey Regional Cancer Registry. Population data were obtained from OPCS and the Registrar General's Office. Smoking data up to 1988 were obtained from Wald and Nicolaidis-Bouman (1991).

Age-period-cohort analyses, using Poisson regression, show that the trends for lung cancer incidence are determined by birth cohort-effects and for males by calendar-period effects as well. Therefore, models using age-standardised or broad-band age-specific rates are not adequate. Incorporation of the average tar content of cigarettes as period effects, lagged by 10 years, and either the number of cigarettes smoked per smoker or the percentage of males/females who smoke as cohort effects in the age-period-cohort models allow investigation of changing patterns in smoking behaviour.

Lung cancer incidence is projected to decrease for all ages in males, and to increase overall in females. However, for younger females incidence will decline. If the decline in the tar content of cigarettes were to continue and either the present trends in smoking continue or the targets for smoking are met, males, but not females will meet the Health of the Nation targets for lung cancer incidence. If the tar content remains at the 1986 level, the targets for lung cancer will not be met for either sex. Assuming no change in treatment patterns from the 1980s, in 2001 one full time equivalent consultant in medical oncology and one full time equivalent consultant radiotherapist will be required to treat lung cancer patients alone, and approximately 200 patients will require surgery.

This thesis has demonstrated a useful methodology for incorporating risk factor information in projection models for lung cancer incidence, that could be generalized to projections for other registries.



## ACKNOWLEDGEMENTS

The original ideas for this thesis were conceived when I was working at the Ontario Cancer Registry, Canada. I would like to thank Dr Lourraine Marrett and Dr Aileen Clarke for their inspiration and support over the years.

During our years in Kenya, Dr Jane Carter, encouraged me to persevere. I would like to thank her.

On arrival in Liverpool, Dr Hugh Annett rekindled my resolve and I am grateful for the support provided by him and other members of staff of the International Community Health Department, Liverpool School of Tropical Medicine.

This thesis would not have been completed without the guidance and encouragement of my two supervisors, Dr Deborah Ashby and Dr Dave Haran. I greatly appreciate the sympathetic approach they have both taken.

General wisdom states that to complete a thesis, one should neither move geographically or change jobs. I have done both, but am very fortunate that on starting my latest position, Professor Maggie Pearson and Professor Tom Walley put a high priority on the completion of this thesis. I would like to thank them both.

I would also like to thank Professor Sam Leinster and his team at the Royal Liverpool University Hospital for their excellent care and support, without which this thesis would not have been completed.

I also express my appreciation to the many good friends and colleagues who have assisted me over the years.

Lastly, I thank my family for the love and encouragement they have provided: both my late parents and my parents-in-law; my children Suzanne and Ian who can hardly remember life without 'Mum writing her PhD' and my husband Peter who has always believed in me.

Thank you to everyone.

## LIST OF FIGURES AND TABLES

### 1. INTRODUCTION

Figure 1.1	Lung Cancer Incidence among Males in England and Wales, 1935-1968.	6
------------	--	---

### 2. LITERATURE REVIEW

Figure 2.1	Cancer Incidence Rates and Fitted Line, Ontario, 1969-1994, All Sites, Males aged 45-64 years	32
Figure 2.2	Age-Specific Lung Cancer Mortality Rates among Men and Women in Spain, 1951-80	33
Figure 2.3	Age-Specific Lung Cancer Mortality Rates, 1951-1980, among Men and Women in Spain for Selected Birth Cohorts	34
Figure 2.4	Mortality Rates for Cancer of the Stomach in Females in Denmark, 1943-1978	35
Figure 2.5	Lung Cancer Mortality among Males in England and Wales, 1935-1968, by Time Period and Year of Birth. From Doll (1971)	39
Figure 2.6	Age-Specific Lung Cancer Mortality Rates in Belgium, 1955-1978 Female	39
Table 2.1	Hierarchy of Model Fitting	42
Figure 2.7	Mortality Rates from Lung Cancer in Women aged 40-44 in England and Wales during 1951-80. Estimated Rates for 1971-80 based upon 1951-70 are also plotted for the Two Methods	44
Figure 2.8a	Relative Risk for Lung Cancer according to the Number of Cigarettes Smoked per Day, Male	60
Figure 2.8b	Relative Risk for Lung Cancer according to the Number of Cigarettes Smoked per Day, Female	60
Figure 2.9	Relative Risk for Lung Cancer according to the Number of Cigarettes Smoked per Day, and Duration, Male	61



Figure 2.10	Lung Cancer Risk according to Lifetime Cigarette Consumption, Canada 1981-1985	61
Figure 2.11a	Relative Risk for Lung Cancer according to the Time since Quitting Male	63
Figure 2.11b	Relative Risk for Lung Cancer according to the Time since Quitting Female	63
Figure 2.12	Relative Risk of Lung Cancer by Time since Quit and Duration of Smoking Male	64
Figure 2.13	Mortality Ratios for Lung Cancer according to Age at Starting to Smoke Male	64
<b>3. INFORMATION REQUIREMENTS FOR MODELLING LUNG CANCER INCIDENCE RATES</b>		
Table 3.1	Administrative Areas in LRHB	82
Figure 3.1	Map of Cheshire and Lancashire showing LRHB and MRHA	83
Table 3.2	Health Districts in MRHA	84
Table 3.3	Definition of Geographic Area	85
Table 3.4	Information Requested on Each Cancer Case	86
Figure 3.2	Percent of Registered Lung Cancer Cases Histologically Confirmed, by Sex	90
Table 3.5	Summary of Smoking Data Available	100
Table 3.6	Information Sources for Data relating to Lung Cancer Incidence, Population and Service Needs	103
Table 3.7	Information Source for Data on Risk Factors for Lung Cancer	104
<b>4. TRENDS IN LUNG CANCER INCIDENCE IN MERSEY REGION AND SIMPLE METHODS OF PROJECTIONS</b>		
Figure 4.1	Number of Incident Lung Cancer Cases, by Sex	109

Figure 4.2	Crude Lung Cancer Incidence Rates, by Sex and Year	112
Table 4.1	Age Specific Rates per 100,000 for Male Lung Cancer for LRHB (1959-73) and MRHA (1974-88)	115
Table 4.2	Age Specific Rates per 100,000 for Female Lung Cancer for LRHB (1959-73) and MRHA (1974-88)	116
Figure 4.3	Age-Standardised Lung Cancer Incidence Rates, by Sex	119
Figure 4.4	Broad-Band Age-Specific Lung Cancer Incidence Rates, Male	122
Figure 4.5	Broad-Band Age-Specific Lung Cancer Incidence Rates, Female	122
Figure 4.6	Age-Specific Lung Cancer Incidence Rate by Quinquennia, Male	125
Figure 4.7	Age-Specific Lung Cancer Incidence Rate by Quinquennia, Female	126
Figure 4.8	Age-Specific Lung Cancer Incidence Rate by Birth Cohort, Male	128
Figure 4.9	Age-Specific Lung Cancer Incidence Rate by Birth Cohort, Female	129
Figure 4.10	Projected Age-Standardised Lung Cancer Incidence Rates, Male	134
Figure 4.11	Projected Age-Standardised Lung Cancer Incidence Rates, Female	135
Figure 4.12	Projected Broad-Band Age-Specific Cancer Incidence Rates, Male	141
Figure 4.13	Projected Broad-Band Age-Specific Cancer Incidence Rates, Female	142
 <b>5. TRENDS IN SMOKING BEHAVIOUR</b>		
Figure 5.1	Manufactured Cigarettes Smoked per Person per Year, by Sex, UK	153
Figure 5.2	Manufactured Cigarettes Smoked per Person per Week, by Sex and Age, GB Sales Adjusted	155
Figure 5.3	Percentage who Smoke Manufactured Cigarettes, by Sex, GB Sales Adjusted	156



Figure 5.4	Percentage who Smoke Manufactured Cigarettes, by Sex and Age, GB Sales Adjusted	158
Figure 5.5	Manufactured Cigarettes Smoked per Smoker per Week, by Sex, GB Sales Adjusted	159
Figure 5.6	Manufactured Cigarettes Smoked per Smoker per Week, by Sex and Age, GB Sales Adjusted	160
Figure 5.7	Percentage of Smokers who Smoke Plain Cigarettes, by Sex, GB	162
Figure 5.8	Average Tar Yield per Cigarette, UK Sales Adjusted	163
6.	<b>MODELS BASED ON AGE-STANDARDISED LUNG CANCER INCIDENCE AND INFORMATION ON SMOKING BEHAVIOUR</b>	
Table 6.1	Variables Used in the Regression	170
Table 6.2	Data Used in the Regression Analysis	171
Table 6.3	Results of Fitting Male Age Standardised Rates to Smoking Variables	174
Table 6.4	Results of Fitting Female Age Standardised Rates to Smoking Variables	176
Figure 6.1	Projected Age-Standardised Male Lung Cancer Incidence Rate, using Models incorporating Smoking Information	182
Figure 6.2	Projected Age-Standardised Female Lung Cancer Incidence Rate, using Models incorporating Smoking Information	183
Table 6.5	Projected Service Needs for Incident Lung Cancer in Mersey Region based on Projected Age-Standardised Rates, and Information on Smoking Behaviour	185
7.	<b>MODELS BASED ON BROAD-BAND AGE-SPECIFIC LUNG CANCER INCIDENCE RATES AND INFORMATION ON SMOKING BEHAVIOUR</b>	
Table 7.1	Results of Fitting Broad-Band Age-Specific Lung Cancer Incidence Rates to Smoking Variables for Males aged 65+ Years	196
Table 7.2	Results of Fitting Broad-Band Age-Specific Lung Cancer Incidence Rates to Smoking Variables for Males aged 45-64 Years	197

Table 7.3	Results of Fitting Broad-Band Age-Specific Lung Cancer Incidence Rates to Smoking Variables for Females aged 65+ Years	199
Table 7.4	Results of Fitting Broad-Band Age-Specific Lung Cancer Incidence Rates to Smoking Variables for Females aged 45-64 Years	200
Table 7.5	The 'Best' Models for each Sex and Age Group	202
Figure 7.1	Projected Male Broad-Band Age-Specific Lung Cancer Incidence Rates, Models incorporating Smoking Information	204
Figure 7.2	Projected Female Broad-Band Age-Specific Lung Cancer Incidence Rates, Models incorporating Smoking Information	205
Table 7.6	Service Need Projected for 2000 using Broad-Band Age-Specific Incidence Rates and Smoking Information	207
<b>8. AGE-PERIOD-COHORT MODELLING</b>		
Table 8.1	Hierarchy of Model Fitting	227
Table 8.2a	Male Lung Cancer Incidence Frequencies, LRHB 1959-1973, MRHA 1974-1988	232
Table 8.2b	Male Population Frequencies (1,000s) LRHB 1959-1973, MRHA 1974-1988	232
Table 8.3	Parameter Estimates for Male Lung Cancer from the Age-Period-Cohort Modelling	235
Table 8.4	Second Order Differences in Parameter Estimates for Male Lung Cancer from the Age-Period-Cohort Modelling	236
Figure 8.1a	Age Parameter Estimates Male Lung Cancer Age-Period-Cohort Analysis, 1959-1988 LRHB 1959-1973, MRHA 1974-1988	237
Figure 8.1b	Period Parameter Estimates Male Lung Cancer Age-Period-Cohort Analysis, 1959-1988 LRHB 1959-1973, MRHA 1974-1988	237



Figure 8.1c	Cohort Parameter Estimates Male Lung Cancer Age-Period-Cohort Analysis, 1959-1988 LRHB 1959-1973, MRHA 1974-1988	237
Table 8.5a	Female Lung Cancer Incidence Frequencies, LRHB 1959-1973, MRHA 1974-1988	239
Table 8.5b	Female Population Frequencies (1,000s) LRHB 1959-1973, MRHA 1974-1988	239
Table 8.6	Parameter Estimates for Female Lung Cancer from the Age-Period-Cohort Modelling	241
Table 8.7	First Order Differences in Parameter Estimates for Female Lung Cancer from the Age-Period-Cohort Modelling	242
Figure 8.2a	Age Parameter Estimates Female Lung Cancer, Age-Period-Cohort Analysis, 1959-1988, LRHB 1959-1973, MRHA 1974-1988	243
Figure 8.2b	Cohort Parameter Estimates Female Lung Cancer Age-Period-Cohort Analysis, 1959-1988 LRHB 1959-1973, MRHA 1974-1988	243
Table 8.8	Age Groups, Period and Cohorts	247
Table 8.9	Estimated Period and Cohort Parameters for Lung Cancer, MRHA	250
Figure 8.3	Projected Age-Specific Rates, Age-Period-Cohort Model, Male	252
Figure 8.3	Projected Age-Specific Rates, Age-Period-Cohort Model, Female	253
Figure 8.5	Comparison of Observed and Projected Annual Numbers of Cancer Cases, 1991 Quinquennia, Male	254
Figure 8.6	Comparison of Observed and Projected Annual Numbers of Cancer Cases, 1991 Quinquennia, Female	254
Table 8.10	Projected Service Needs for the Year 2000	256

9. MODELS BASED ON AGE-PERIOD-COHORT ANALYSIS  
INCORPORATING INFORMATION ON SMOKING BEHAVIOUR

Table 9.1	Relation between Periods and Cohorts for Age-Specific Smoking Information	263
Table 9.2	Changes in Scaled Deviance for the Different Lag Times for Average Tar Content	265
Table 9.3	Parameter Estimates for Male Lung Cancer from the Age-Cohort Average Tar Model (Model 5a)	266
Table 9.4	Correspondence between Cohorts for Lung Cancer Incidence and the Smoking Variables	268
Table 9.5	Residual Scaled Deviances for the Model relating Smoking Variable to Male Lung Cancer Incidence	270
Figure 9.1a	Age-Period-Cohort Analysis, Age Parameter Estimates, Male	271
Figure 9.1b	Age-Period-Cohort Analysis, Cohort Parameter Estimates, Male	272
Table 9.6	Parameter Estimates for Male Lung Cancer from the Age, Percentage of Males who Smoke and Average Tar Modelling (Model 6a)	273
Table 9.7	Parameter Estimates for Male Lung Cancer from the Age, Number of Cigarettes Smoked per Male and Average Tar Modelling (Model 6b)	274
Table 9.8	Residual Scaled Deviances for the Model relating Smoking Variable to Female Lung Cancer Incidence	275
Table 9.9	Parameter Estimates for Female Lung Cancer from the Age, Percentage of Females who Smoke and Average Tar Modelling (Model 6a)	276
Figure 9.2a	Age-Period-Cohort Analyses, Age Parameter Estimates, Female	277
Figure 9.2b	Age-Period-Cohort Analyses, Cohort Parameter Estimates, Female	277

Table 9.10	Parameter Estimates for Female Lung Cancer from the Age, Number of Cigarettes Smoked per Female and Average Tar Modelling (Model 6b)	278
Figure 9.3a	Age Parameters Estimates for Model 6 including % who Smoke, Males and Females	280
Figure 9.3b	Cohort Parameters Estimates for Model 6 including % who Smoke, Males and Females	280
Table 9.11	The Scenarios used for Projections	284
Table 9.12	Estimates of the Average Tar Content of Cigarettes	286
Table 9.13	Estimates of the Smoking Variables for Males	288
Table 9.14	Estimates of the Smoking Variables for Females	289
Figure 9.4	Age-Standardised Lung Cancer Incidence Rates, Male	292
Figure 9.5a	Projected Lung Cancer Incidence Rates, Scenario 9.1, Male	293
Figure 9.5b	Projected Lung Cancer Incidence Rates, Scenario 9.3, Male	293
Figure 9.5c	Projected Lung Cancer Incidence Rates, Scenario 9.5, Male	294
Figure 9.5d	Projected Lung Cancer Incidence Rates, Scenario 9.7, Male	294
Figure 9.6	Age-Standardised Lung Cancer Incidence Rates, Female	295
Figure 9.7a	Projected Lung Cancer Incidence Rates, Scenario 9.1, Female	296
Figure 9.7b	Projected Lung Cancer Incidence Rates, Scenario 9.3, Female	296
Figure 9.7c	Projected Lung Cancer Incidence Rates, Scenario 9.5, Female	297
Figure 9.7d	Projected Lung Cancer Incidence Rates, Scenario 9.7, Female	297



Figure 9.8	Comparison of Observed and Projected Annual Numbers of Cancer Cases, 1991 Quinquennia, Male	299
Figure 9.9	Comparison of Observed and Projected Annual Numbers of Cancer Cases, 1991 Quinquennia, Female	300
Table 9.15	Projected Percentage Reduction in Male Lung Cancer 1991 to 2011 for Males aged less than 75 years	301
Table 9.16	Projected Percentage Reduction in Female Lung Cancer 1991 to 2011 for Females aged less than 75 years	302
Tables 9.17	Projected Service Needs for Incident Male Lung Cancer in Mersey Region	304
Tables 9.18	Projected Service Needs for Incident Female Lung Cancer in Mersey Region	305

## 10. DISCUSSION

Table 10.1	Percentage of the Population who Smoke: A Comparison of the MRHA and GB	317
------------	---	-----

## LIST OF APPENDICES

### APPENDIX III

1	Fifth Digit Morphology Codes	373
2a	Residence Codes	374
2b	Postal Codes	376
3a	Treatments Received by Male Lung Cancer Patients in Mersey Region, 1983-87	377
3b	Treatments Received by Female Lung Cancer Patients in Mersey Region, 1983-87	378

### APPENDIX IV

1	Age-standardised Lung Cancer Incidence Rates, by sex,	379
2	Age-Specific Lung Cancer Incidence Rates by Quinquennia, Male	379
3a	Projected Male Lung Cancer Incidence. Linear Projections of Log Age-Standardised Rates against Year	380
3b	Projected Female Lung Cancer Incidence. Linear Projections of Log Age-Standardised Rates against Year	381
4a	Projected Male Lung Cancer Incidence. Linear Projections of Broad-Band Age-Specific Rates against Year	382
3b	Projected Female Lung Cancer Incidence. Linear Projections of Broad-Band Age-Specific Rates against Year	385

### APPENDIX VI

	Projected Age-Standardised Lung Cancer Incidence Rates for Mersey Region using Models incorporating Smoking Information	388
--	---	-----

### APPENDIX VII

	Projected Broad-Band Age-Specific Lung Cancer Incidence Rates for Mersey Region using Models incorporating Smoking Information	389
--	--	-----

### APPENDIX VIII

a	Projections for Male Lung Cancer Incidence in MRHA based on Age-Period-Cohort Analysis	390
---	--	-----

b	Projections for Female Lung Cancer Incidence in MRHA based on Age-Period-Cohort Analysis	391
---	--	-----

#### APPENDIX IX

1a	Scenario 9.1: Projections for Male Lung Cancer Incidence in MRHA based on Extrapolation of Average Tar Content and Percentage of Males who Smoke	392
1b	Scenario 9.1: Projections for Female Lung Cancer Incidence in MRHA based on Extrapolation of Average Tar Content and Percentage of Females who Smoke	393
2a	Scenario 9.2: Projections for Male Lung Cancer Incidence in MRHA based on Extrapolation of Average Tar Content and the Health of the Nation target for the Percentage of Males who Smoke	394
2b	Scenario 9.2: Projections for Female Lung Cancer Incidence in MRHA based on Extrapolation of Average Tar Content and the Health of the Nation target for the Percentage of Females who Smoke	395
3a	Scenario 9.3: Projections for Male Lung Cancer Incidence in MRHA based on 1986 Average Tar Content and Percentage of Males who Smoke	396
3b	Scenario 9.3: Projections for Female Lung Cancer Incidence in MRHA based on 1986 Average Tar Content and Percentage of Females who Smoke	397
4a	Scenario 9.4: Projections for Male Lung Cancer Incidence in MRHA based on 1986 Average Tar Content and the Health of the Nation target for the Percentage of Males who Smoke	398
4b	Scenario 9.4: Projections for Female Lung Cancer Incidence in MRHA based on Extrapolation of Average Tar Content and the Health of the Nation target for the Percentage of Females who Smoke	399
5a	Scenario 9.5: Projections for Male Lung Cancer Incidence in MRHA based on Extrapolation of Average Tar Content and Number of Cigarettes Smoked per Male	400
5b	Scenario 9.5: Projections for Female Lung Cancer Incidence in MRHA based on Extrapolation of Average Tar Content and Number of Cigarettes Smoked per Female	401



6a	Scenario 9.6: Projections for Male Lung Cancer Incidence in MRHA based on Extrapolation of Average Tar Content and the Health of the Nation target for the Number of Cigarettes Smoked per Male	402
6b	Scenario 9.6: Projections for Female Lung Cancer Incidence in MRHA based on Extrapolation of Average Tar Content and the Health of the Nation target for the Number of Cigarettes Smoked per Female	403
7a	Scenario 9.7: Projections for Male Lung Cancer Incidence in MRHA based on 1986 Average Tar Content and Number of Cigarettes Smoked per Male	404
7b	Scenario 9.7: Projections for Female Lung Cancer Incidence in MRHA based on 1986 Average Tar Content and Number of Cigarettes Smoked per Female	405
8a	Scenario 9.8: Projections for Male Lung Cancer Incidence in MRHA based on 1986 Average Tar Content and the Health of the Nation target for the Number of Cigarettes Smoked per Male	406
8b	Scenario 9.8: Projections for Female Lung Cancer Incidence in MRHA based on 1986 Average Tar Content and the Health of the Nation target for the Number of Cigarettes Smoked per Female	407

Chapter 1

**INTRODUCTION**



This study was undertaken to investigate the trends in lung cancer incidence in Mersey Region and to develop models for the projection of lung cancer incidence. The models that are investigated are based on past trends of lung cancer incidence in Mersey Region. The relationships of these trends to past trends in risk behaviour will also be investigated. Understanding of the trends and relationships should lead to improved projections of lung cancer incidence in Mersey Region. The projections are used to determine whether the Health of the Nation targets for the reduction in lung cancer will be met, and the future service needs for lung cancer patients in Mersey Region. The projections will also be useful in answering other questions relating to future lung cancer incidence.

This chapter will:

- 
- briefly describe the epidemiology of lung cancer in Mersey Region
  - give an overview of the methods of projection that will be used in this thesis
  - describe the Health of the Nation strategy
  - describe the background for the determination of service need for lung cancer patients
  - outline the chapters contained in the thesis
-

## 1.1 Lung Cancer Incidence

In the period 1986-90 lung cancer was the most frequently occurring cancer among males in Mersey Region, accounting for 24.3% of all cancers (Williams et al 1993). In females it accounted for 12.2%, and only breast cancer (20.3% of all cancers) and non melanoma skin cancer (14.8% of all cancers) had higher incidence. The crude incidence rates in 1990 were 125.8 per 100,000 for males and 64.6 per 100,000 for females.

Data on lung cancer incidence is available from the Mersey Regional Cancer Registry. This registry initiated registration of incident cancer cases in Mersey Region and North Wales in 1944 (personal communication Sandra Gravestock). Records have been computerised since 1951. For each incident lung cancer case registered available information includes diagnosis, morphology, date of diagnosis, age, sex and area of residence. Subsequent to the reorganisation of Regional Health Authorities in 1994, Mersey Regional Health Authority has become part of North West Regional Health Authority, and the Mersey Regional Cancer Registry has been renamed the Merseyside and Cheshire Regional Cancer Registry. In this thesis the previous nomenclature will be used.

This thesis will:

---

describe the trends in lung cancer incidence in Mersey Region between 1951 and 1988

---

Before the advent of cancer registries, cancer mortality was used to describe the trends in lung cancer. The median survival time in Mersey Region for lung cancer is short (4 months) and therefore trends in incidence would be expected to be similar to trends in lung cancer mortality. Lung cancer mortality rates have risen sharply throughout the majority of the twentieth century. In England and Wales the standardised mortality rates rose from 1.1 per 100,000 for males and 0.7 per 100,000 for females in 1901-20 to 10.6 and 2.5 per 100,000 for males and females respectively in 1936-39 (Stocks 1947). This increase was attributed by some to improved standards of diagnosis (eg Clemmenson and Busk 1947), while others believed that the increase was real (eg Stocks 1947).

The increases presented above relate to increases over calendar time. This would imply that factors affecting lung cancer incidence would affect all age groups in the same fashion at the same time. Another way of comparing rates over time is to investigate the changes over birth

cohorts where:

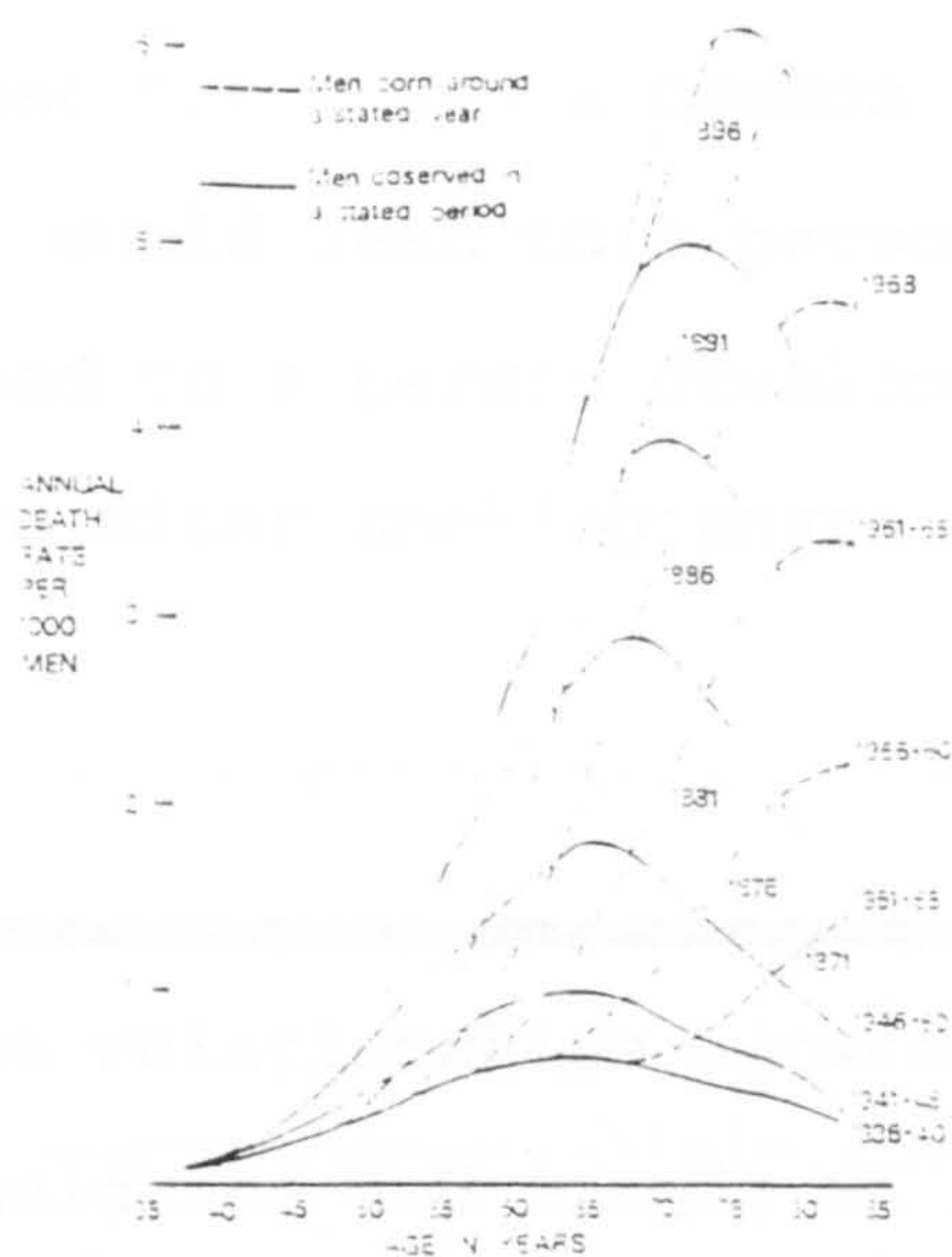
A birth cohort is defined as those people born within a given time period

The difference in trends over calendar period and birth cohort is illustrated in Figure 1.1. The solid lines give the age-specific rates for a given calendar period. Comparison of these curves indicate increasing mortality over calendar period, with a maximum lung cancer mortality around the age of 65 years for any given calendar period. The dashed lines give the age-specific mortality rates for specific cohorts. Comparison of these curves would indicate an increase in mortality over cohorts. The maximum age-specific incidence at age 65 years is not apparent, and is probably due to changes over birth cohort.

The birth cohorts for lung cancer incidence that are investigated in this thesis are usually defined as those people born within a ten year span (Section 7.1) If the changes observed in lung cancer incidence relate to birth cohorts, this would imply that changes in factors that affect lung cancer incidence would affect the birth cohorts differentially.



Figure 1.1 Lung Cancer Incidence among Males in England and Wales, 1935-1968. (Doll 1971)



From Day and Charnay 1982

This thesis will :

---

investigate trends in lung cancer incidence over both calendar period and birth cohort

---

Two main causes for lung cancer were proposed, general atmospheric pollution and tobacco smoking (Doll and Hill 1950). The levels of atmospheric pollution from car exhaust, from gas works, from industrial plants and coal fires did increase in the beginning of the century, as did the levels of smoking, among both males and females. Doll and Hill (1950) demonstrated a strong relationship between smoking and lung cancer in both males and females. They argued against lung cancer causing people

to smoke, because the smoking habit was usually developed before the onset of the cancer. They also did not support the hypothesis that there was a common cause, unspecified, that would lead to a person beginning to smoke and also lead to a person developing lung cancer, usually many years after smoking began.

This thesis will:

---

investigate the relationships between trends in lung cancer incidence and trends in smoking behaviour

---

## 1.2 Projections

The WHO/IARC Expert Committee (1979) defined projections as:

a method of describing the implications of certain assumptions about future trends without necessarily attaching any measure of likelihood to them.

If it is possible to determine which of the assumptions are the most likely to happen, the projection based on these assumptions would be defined as a **prediction**.

In order to investigate future lung cancer incidence, it is important to understand the past trends in incidence.

This may be achieved by visual examination of the plots of the incidence rates over time. However, it is preferable to be able to quantify the trends, so that mathematical models can be developed to assist in the projections.

This thesis will:

---

develop models to describe trends in lung cancer incidence

---

If the trends in lung cancer incidence are approximately linear, the simplest form of projections is to extend the trend to the years of interest, ie linear extrapolation. This could be achieved by using a ruler. However, it is preferred to use regression methods to relate some form of the incidence rates to the year of diagnosis. The models determined from these methods can then be used for projecting lung cancer incidence.

This thesis will:

---

use the models to describe trends in lung cancer incidence to project future lung cancer incidence by linear extrapolation over calendar time

---

Projections based on these methods assume that the past trends will continue for the period of projection.



Improvements in projections may be achieved by including information on risk factors, such as smoking behaviour. If a relationship between the trends in lung cancer incidence and smoking behaviour can be determined then future lung cancer incidence can be determined from smoking behaviour.

This thesis will:

---

project future lung cancer incidence using models incorporating information on smoking behaviour

---

If the lag time between the smoking behaviour and the cancer incidence is of long enough duration, the future cancer incidence can be modelled on past smoking behaviour

This thesis will:

---

determine the appropriate lag time between trends in smoking behaviour and trends in lung cancer incidence

---



### 1.3 Health of the Nation

In 1992 the British government developed the Health of the Nation strategy (Department of Health 1993). The strategy aims to improve the health of the English people in line with the goals inherent in the WHO's goal of 'Health for All' by the year 2000. To achieve these goals the concept of health promotion, as well as health care is included in the mandate of the National Health Service.

The Health of the Nation strategy targets certain diseases and mechanisms have been identified to reduce the incidence and/or mortality of these diseases. Cancer, and in particular lung cancer, is one of the targeted disease.

The target for the reduction of lung cancer is:

#### **Health of the Nation Targets for Lung Cancer**

To reduce the death rate for lung cancer under the age of 75 by at least 30% in men and by at least 15% in women by 2010 (from 60 per 100,000 for men and 24.1 per 100,000 for women in 1990 to no more than 42 and 20.5 respectively)

The mechanism to reduce the burden of mortality of this disease is to reduce smoking among the English population. In particular four major targets have been identified. The first two relate to the smoking behaviour

of the population in general:

1. To reduce the prevalence of cigarette smoking to no more than 20% by the year 2000 in both men and women (a reduction of at least 35% in man and 29% in women, from a prevalence in 1990 of 31% and 28% respectively)
2. To reduce consumption of cigarettes by at least 40% by the year 2000 (from 98 billion manufactured cigarettes per year in 1990 to 59 bn)

The second two relate to subsections of the population:

3. In addition to the overall reduction in prevalence, at least 33% of women smokers to stop smoking at the start of their pregnancy by the year 2000
4. To reduce smoking prevalence of 11-15 year olds by at least 33% by 1994 (from about 8% in 1988 to less than 6%)

This thesis will:

---

investigate whether the Health of the Nation targets are projected to be attained, for each set of projections

investigate the effect of achieving the Health of the Nation targets for the percentage of the population who smoke on lung cancer incidence

investigate the effect of achieving the Health of the Nation target for the quantity smoked

---

The two other targets for smoking are not investigated in this thesis.

#### 1.4 Service Needs

Approximately 90 hospital beds in Mersey Region are occupied by lung cancer patients every day (Williams et al 1993). It is estimated that in 1990/91 the annual hospital cost of lung cancer in Mersey Region was around £5.25 million. This figure excludes non-hospital costs which are considerable.

The Chief Medical Officer's Expert Advisory Group on Cancer (Department of Health 1994) has developed a framework for the rationalisation of cancer services within England. In order to implement the recommendations of this report information on the number of cancer patients that can be expected in the future is required. Information on expected treatment patterns are also required. In this thesis these treatment patterns are assumed to remain the same as the treatment patterns for the mid 1980s. No more recent information is available.

This thesis will:

---

determine the service needs for lung cancer patients at the beginning of the 21st century

---

## 1.5 Outline of Thesis

This thesis consists of ten chapters. In each chapter where projected lung cancer incidence is calculated, the projections for 1990 or 1991 are compared to those actually observed. The changes in lung cancer incidence between 1990 and 2010 are compared to the targets for the Health of the Nation. Service needs for the beginning of the 21st century are calculated based on statistics for past treatment practises. A brief outline of the chapters is presented below:

Chapter 1 introduces the thesis

Chapter 2 presents the review of the literature.

Chapter 3 discusses the data needed for the analyses

Chapter 4 presents the trends in lung cancer incidence and simple methods of projection. These methods are based on linear extrapolation of past trends in lung cancer incidence over calendar period

Chapter 5 presents the trends in smoking behaviour.

Chapter 6 presents methods for including information on smoking behaviour in the projection models based on age-standardised lung cancer incidence rates.

Chapter 7 present methods for including information on smoking behaviour in the projection models based on age-specific lung cancer incidence



rates.

- Chapter 8 explains age-period-cohort analysis. Age-period-cohort models are developed and used for the projection of lung cancer incidence.
- Chapter 9 investigates the incorporation of information on smoking behaviour in the age-period-cohort models. Future lung cancer incidence rates are projected using these models
- Chapter 10 discusses the results, with particular reference to the targets for the Health of the Nation (Department of Health 1993) and the Report of the Chief Medical Officer's Expert Advisory Group on Cancer (Department of Health 1994)

Chapter 2

**LITERATURE REVIEW**

Throughout time the human race has been interested in being able to predict the future. In the field of health care there is increasing interest in investigating future disease burdens. The WHO/IARC Expert Committee (1979) have differentiated two methods for such investigations:

- Projection: a method of describing the implications of certain assumptions about future trends without necessarily attaching any measure of likelihood to them.
- Prediction: a projection that is most likely to represent the future course of events.

It is essential to be able to predict future disease incidence and prevalence to effectively plan preventive, diagnostic and therapeutic services for a given disease (Hakulinen and Pukkula 1981). Mersey Region has one of the highest incidence of lung cancer in England. In 1991, with over 2,100 incident lung cancer cases, it was estimated that lung cancer patients accounted for approximately 33,000 hospital bed days annually at an annual cost of around £5.25 million (Williams et al 1993). The lung cancer incidence rates are increasing for females, although they are declining slowly for males. Given the high cost of treatment, the financial restraints on Health Services, and the recommendations of the Chief Medical Officer's Expert Advisory Group on

Cancer (Department of Health 1994) accurate predictions of lung cancer incidence would greatly assist in planning utilisation of services in Mersey Region.

Projections can be based on the extrapolation of past trends in disease rates alone, or include consideration of trends in risk factors for the disease under consideration (Hakama and Pukkala, 1984, Hakama, Hakulinen and Laara, 1986). For lung cancer a major risk factor is known to be smoking (Doll and Hill 1950).

Inclusion of information on smoking behaviour in the projection models for lung cancer incidence may increase the accuracy of these models. Such models would be useful to facilitate planning and monitoring of intervention programs to reduce lung cancer through the reduction of smoking. The Health of the Nation strategy has set targets for the reduction of cigarette consumption and have hypothesised a corresponding reduction in lung cancer incidence (Department of Health 1993).

Similarly projections could also assist in the evaluation of disease screening programs, through the comparison of the projected incidence and that obtained after the introduction of the screening program. In screening for breast cancer which is aimed at detecting early stage cancers, there would be an initial increase in incidence rates as early stage cancers were detected. As the program continued, the incidence should return to the



projected levels. When screening for cervical cancer, the aim is to detect pre invasive cancers. After initiation of a screening program for this cancer the incidence rates should be lower than those projected on past trends. However, since there is no screening program for lung cancer this aspect will not be considered further.

This literature review will be structured as follows:

- 
- The Uses of Projections
  - The Methodology for Projections
    - Statistical Modelling
    - Time Series including extrapolation of trends over time of:
      - Age-Standardised Rates
      - Age-Specific Rates
      - Age-Period-Cohort Analysis
    - Time Series with the Inclusion of Risk Factors for:
      - Age-Standardised Rates
      - Age-Period-Cohort Analysis
  - Aetiology of Lung Cancer
    - Cigarette Smoking
    - Occupational Exposure
    - Air Pollution
    - Diet
    - Social Class
  - Effects of Interventions
-

## 2.1 Uses of Projections

Projections of cancer incidence have been used to assist in the planning process for the provision of cancer treatment services. In 1985, the Ontario Cancer Foundation was requested by the Provincial Government to estimate the number of radiotherapy machines that would be required in the province by 1996. The Foundation was also requested to evaluate whether a new Cancer Treatment Centre was justified in North Eastern Ontario (Ontario Cancer Treatment and Research Foundation, 1986). The required number of radiotherapy machines was estimated using projections of the number of incident cancer cases that would occur in 1996, together with information on average treatment regimes for cancer patients, and machine capacity. The conclusion of this report was that given the projected increase in the number of cancer patients, the new treatment centre was indeed justified.

In Britain the recommendations of the Chief Medical Officer's Expert Advisory Group on Cancer (Department of Health 1994) give guidelines for the organisation of cancer services. These include minimum patient throughput to maintain specialist expertise. The Thames Cancer Registry has been active in providing estimates of predicted cancer incidence rates to District Health Authorities in the North Thames and South Thames Regional Health Authorities (personal communication Paula Bland).

In Hertfordshire, estimates of predicted cancer incidence rates were helpful in evaluating future service needs and deciding on the need for a new specialist cancer centre. The trends in cancer incidence were such that the projected increase was not large enough to justify a new centre, given the Cancer Centres currently serving the area (personal communication Paul Cosford).

Cancer incidence projections for the ten most common cancer are being used to assist in the rationalisation of cancer service provision in North Essex (personal communication Kevin Loth). The projections assist in evaluating whether the Cancer Units and Cancer Centres providing radiotherapy, chemotherapy and surgery will have enough throughput to maintain their specialist expertise. In East London, a similar exercise is currently being undertaken using the projections to assist in the decision making process (personal communication Dinesh Sethi). In times of fiscal constraint and new policy, projections can assist in the rationalisation of treatment services.

In East Kent projections of cancer incidence are being used to monitor the progress towards the attainment of the Health of the Nation targets (personal communication Harold Elwood)



The value of the predictions in assisting the decision making process depends on the reliability and validity of the actual projections used. This depends in part on the methods used for the projections. The next section presents a review of models that have been used to project cancer incidence or mortality.

## 2.2 Methodology for Projections

There are two main methods for projecting future disease rates:

- Statistical modelling, incorporating the underlying biological model.
- Trend analysis of the time series of rates.

This section will describe projection models arising from these two main methods. Some of the models have been used to describe mortality rates, rather than incidence rates. Mortality data are more readily available than incidence data. The median survival time for lung cancer in Mersey Region is 4 months (Williams et al 1993). Given this short survival time the models for lung cancer mortality and lung cancer incidence can be considered interchangeable.



### 2.2.1 Statistical Modelling

In the first approach, the biological disease process is used to develop an explanatory statistical model. Doll (1971) suggest a model for an individual's lung cancer incidence risk that is based on the number of cigarettes smoked per week and the age at starting smoking, of the form:

$$R_i = b N k (i-w)^{(k-1)}$$

where  $R_i$  is the age-specific rate for age group  $i$

$b$  is a constant

$N$  is the number of cigarettes smoked per week

$w$  is the age of starting to smoke and

$k$  is a constant.

Townsend (1978) adapts this to lung cancer mortality in population cohorts whose smoking behaviour varies over time. The first term in the model relates to the distribution of "durations of smoking" of the cohort at time  $j$  and the average level of consumption at which this smoking took place. Smoking that takes place more recently is weighted more heavily than smoking which took place in the more distant past. The effects of the amount of the tobacco in the cigarette, the tar content and whether the cigarettes are plain or filter are also included in this term. A second term in the model is similar to the first, but relates to non cigarette smoking. The final term relates to the lung cancer

incidence among non-smokers. An assumption is made that the risk for a non-smoker remains similar to the risk at the point of stopping smoking. The data on quantity smoked for the cohorts of interest (1886-90 through 1931-35) is generated from the age-specific data on quantity smoked available from 1948 through 1960 and the total cigarette consumption available from 1870 onwards. Other information on the composition of cigarettes is available from the 1960s.

The overall fit of the model is good for both males and females. The projections for 1973 are compared to the observed rates for the same year. The agreement is good for males aged less than 65 years, but the increase that was observed in the two oldest age groups is not predicted. For women the projections tend to overestimate the rise in mortality for those aged 45 years and above, and do not predict the observed fall in female mortality in the younger age groups.

While the theoretical models attempt to explain the possible variations in smoking behaviour within cohort, a major problem is the lack of adequate smoking data for the estimation of all the smoking variables included. Also some of the assumptions are too simplistic, such as that related to the effect of quitting smoking.

Hakulinen and Pukkala (1981) develop a deterministic

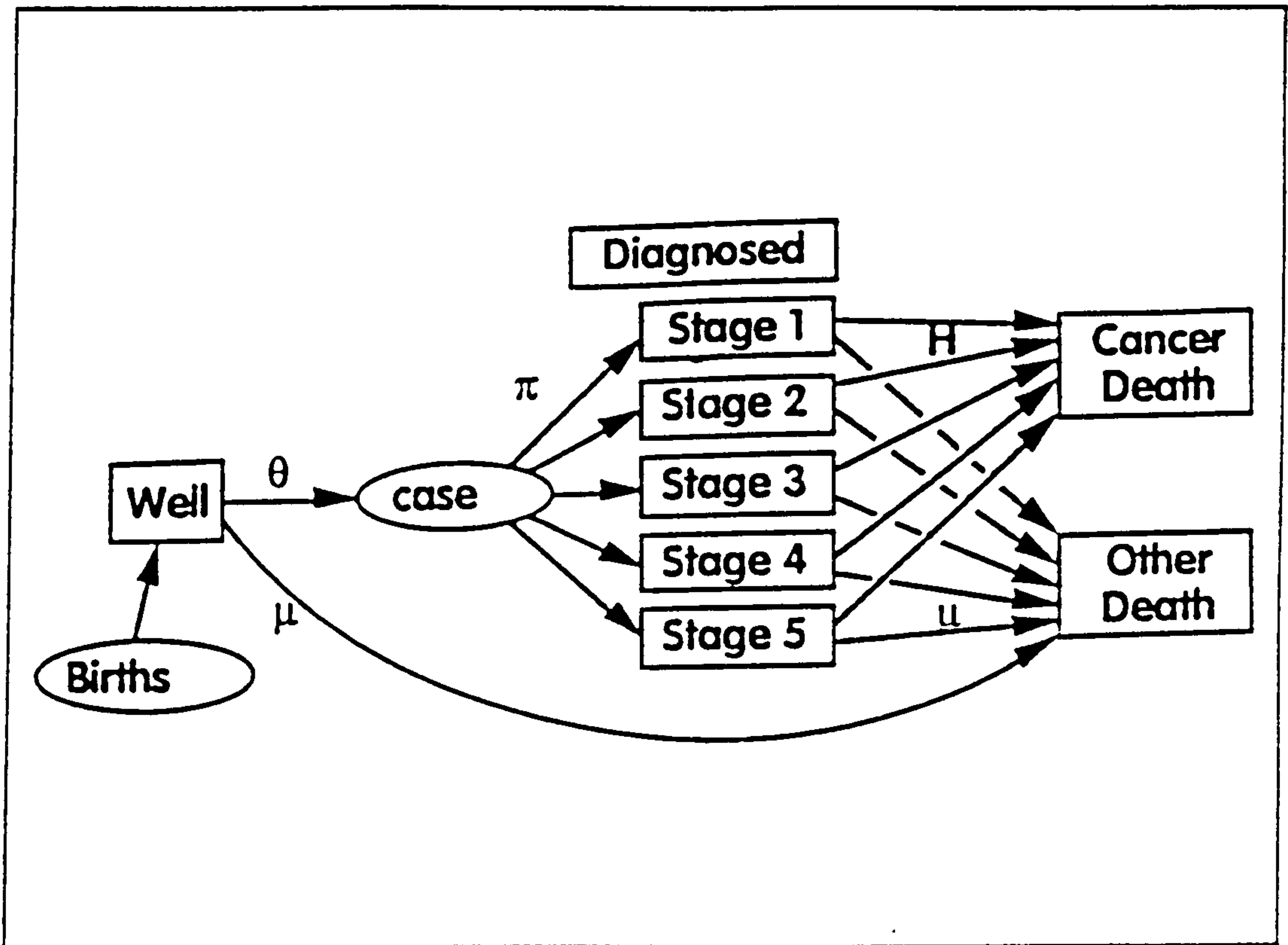
Markov model for lung cancer incidence with 5 year discrete time steps. They use published risk ratios to determine effects of different smoking behaviours on the development of lung cancer. The smoking variables they investigate are:

- the percentage of smokers, ex-smokers and non-smokers in the population,
- the quantity smoked per smoker,
- the duration of smoking and
- the time since quitting for the ex-smokers.

They use a base population whose profile relates to the age distribution, lung cancer incidence and smoking habits of Finland in 1975. The simulation model used enables the effects of changes in smoking behaviours to be investigated. The main results are presented for projections for the year 2050. The major differences in the projections relate to the percentage of the population who smoke, and the number of cigarettes smoked at the initiation of smoking. These differences are not seen by the year 2000. The projection period (75 years) is extremely long and one would expect that the variation around the estimates would be large. However, no confidence intervals are given for the estimates and therefore it is not possible to determine their accuracy.



Levin et al (1986) develop a transition model to describe the natural history of cancer. They assign people to one of four states: WELL, DIAGNOSED, CANCER DEATH and OTHER DEATH.



The category DIAGNOSED can be subdivided into the different stages of disease at which a person is initially diagnosed. Transition probabilities are estimated from data published on:

- cancer incidence, by site, age, sex and stage;
- mortality for both cancer and other causes;
- cancer survival hazard; and
- the population under study.



Once the baseline transition probabilities have been estimated future cancer incidence and/or mortality can be estimated under different assumptions related to the risk factors. This would enable intervention programs to be evaluated, by their effect on these transition probabilities.

The authors suggest that the major benefits of these approaches are that they offer a rational framework for utilising available data on the cancer status of a population, and identifying aspects that require further investigation. It also allows direct estimation of the effects of different intervention programs. The appropriateness of the algorithms used in the model can be a major limitation. They are often simplistic and do not allow variation of basic parameters, such as future relative survival rates. Another limitation can be the lack of available and relevant information on risk factors for a given population. For example; the cohort smoking data, in the model proposed by Townsend (1978), was back generated from recent cohort information assuming that patterns over cohorts had remained stable for over 40 years.

Another approach is to consider the annual incidence rates as a time series. Methods to analyse such a time series are discussed in the next section.

### 2.2.2 Time Series

In the second approach discussed in this literature review, the annual cancer incidence rates are considered to form a time series. An appropriate model can be developed to describe the data. This model is usually based on a time variable, (calendar year and/or birth cohort), and possibly variables relating to risk behaviours. In most projections it is assumed that the model remains stationary for the period under investigation.

The simplest model for cancer incidence is to assume a linear trend over calendar time of some summary measure of the cancer incidence rates. For projections the current trends in incidence are extrapolated and related to population projections (Hakama et al 1986). For some cancer sites (eg prostate and stomach) the trends are linear over birth cohort, but not calendar time (Juel 1983). In this situation linear extrapolation over birth cohort would give more accurate projections.

If the trends in cancer incidence depend on changes over both calendar period and birth cohort, it is possible to estimate parameters for age, period and cohort effects (Barrett 1978, James and Segal 1982, Osmond and Gardner 1982, Holford 1983, 1985, Clayton and Schifflers 1987a, 1987b, Negri et al 1990, Thorn et al 1992, Reissigova et

al 1994). Osmond (1985) proposes extrapolating the period and cohort parameter curves for these models to estimate future age-specific rates.

Projections based on extrapolation of trends in cancer incidence or mortality over time assume that trends in risk behaviours will remain stable, no intervention or screening program has been started, and that there is no change in diagnostic techniques (Hakulinen et al 1986). To overcome the problems due to changes in risk behaviours, models for cancers where the risk factors are known (eg lung and breast) have been developed which incorporate the trends in risk behaviour (Hakama et al 1986, Stevens and Moolgavkar 1979, 1984, Brown and Kessler 1988). Although a longer series will decrease the statistical errors of the projection, the bias may increase as the longer time series may not be sensitive to recent changes in trend (Hakama et al 1986). While the change in population age structure can be accounted for in the above models, there may be problems with the assumptions on which the population projections were based (Hakulinen et al 1986).

The above time series models are based on the assumption that the annual incidence rates are independent of the previous years incidence. However, for models incorporating autoregressive methods or autocorrelated errors to be accurate, the time series must be of



adequate length in order to estimate the error matrix with sufficient accuracy (personal communication, Raj Bhansali). Bhansali felt that series of less than 40 time points would not give accurate enough estimates for projection purposes.

The main advantage of extrapolating past trends in cancer incidence is that it is relatively simple to undertake. The estimates for future cancer burdens do not differ significantly from those produced by time series analysis incorporating autoregressive models. (John McLaughlin, personal communication). For the cancer sites where there have been significant changes in risk behaviours, such as lung cancer, indicators of the risk behaviours can be incorporated into the model.

A workshop held in Canada examined the advantages and disadvantages of the available cancer projection methodologies. The workshop participants reached a consensus that, for the most part, simple trend analysis would produce the most beneficial results for the purposes of the registries in their relation to planning and policy formulation (McLaughlin, Morgan and Mao 1992)

This section of the literature review will therefore concentrate on investigating projections based on past trends in cancer incidence. The simplest models are based on the linear extrapolation of past trends. These models



can possibly be improved with the incorporation of information on risk behaviours. The methods that this section of the literature review will discuss more fully are:

- extrapolation on year of:
  - age-standardised rates
  - age-specific rates by calendar period
  - age specific rates by birth cohort
  - period and cohort parameter estimates derived from age-period-cohort modelling
- inclusion of risk factor information in models based on:
  - age-standardised rates
  - period and cohort parameter estimates derived from age-period-cohort modelling

#### **2.2.2.1 Age-standardised Incidence Rates**

In order to compare incidence rates over time it is necessary that the summary measure is not affected by the variation in the underlying age distribution of the population in question. One method of achieving this is to directly standardise the rate to a given standard population (Section 4.4.1, Fleiss 1981). The standard population used should be relevant to the comparisons that are to be made. If the comparisons are to be made over time within a given region, then the local population for a relevant year would be appropriate. One choice would be to use the population for the year for

which projected incidence/mortality rates are calculated (Teppo et al 1974). Another choice would be a census year during the period for which the incidence data are available (Thorn et al 1992, Vioque and Bolumar 1987). It is also possible to use a World Standard Population (Waterhouse et al 1982), particularly if international comparisons are to be made (Reissigova et al 1994).

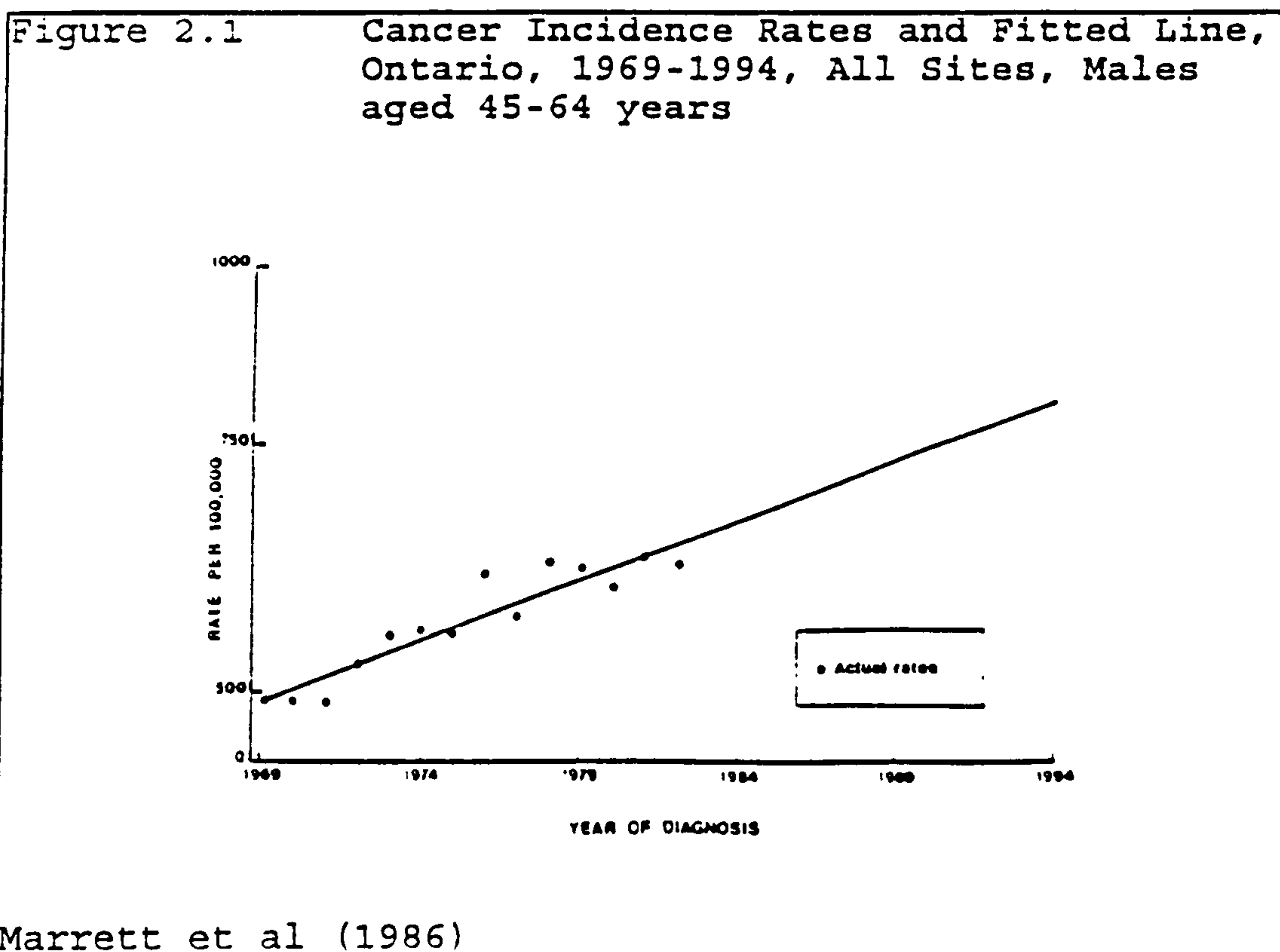
If the age-standardised rates are used the trend in cancer incidence rates over time may be described using either linear or exponential regression analysis.

Although the decision on the form of the regression may be informed by the data, Teppo et al (1974) suggests using linear regression where the trend is increasing and appears linear, and exponential regression if the trend is decreasing. Using the Finnish age-standardised cancer incidence rates for the years 1957 to 1968, Teppo et al (1974) predicts age-standardised incidence rates for 1980 for the main cancer sites. The projected 1980 population for Finland is used as the standard population. Lung cancer incidence amongst males, for example, is projected to increase from 1714 cases in 1968 to between 2000 cases and 2500 cases in 1980 (90% confidence limits). The confidence intervals are calculated using standard regression techniques (Draper and Smith 1981).

The analysis of the trends in age-standardised rates do not allow for the investigation of the potential

differences in trends among the age groups of the age-specific rates. Projections based on age-standardised rates do not allow estimation of age-specific rates or frequencies which may be of importance for planning for service needs where treatment regimens may depend on age (Williams et al 1993), or for evaluation of a screening program aimed at a specific age-group.

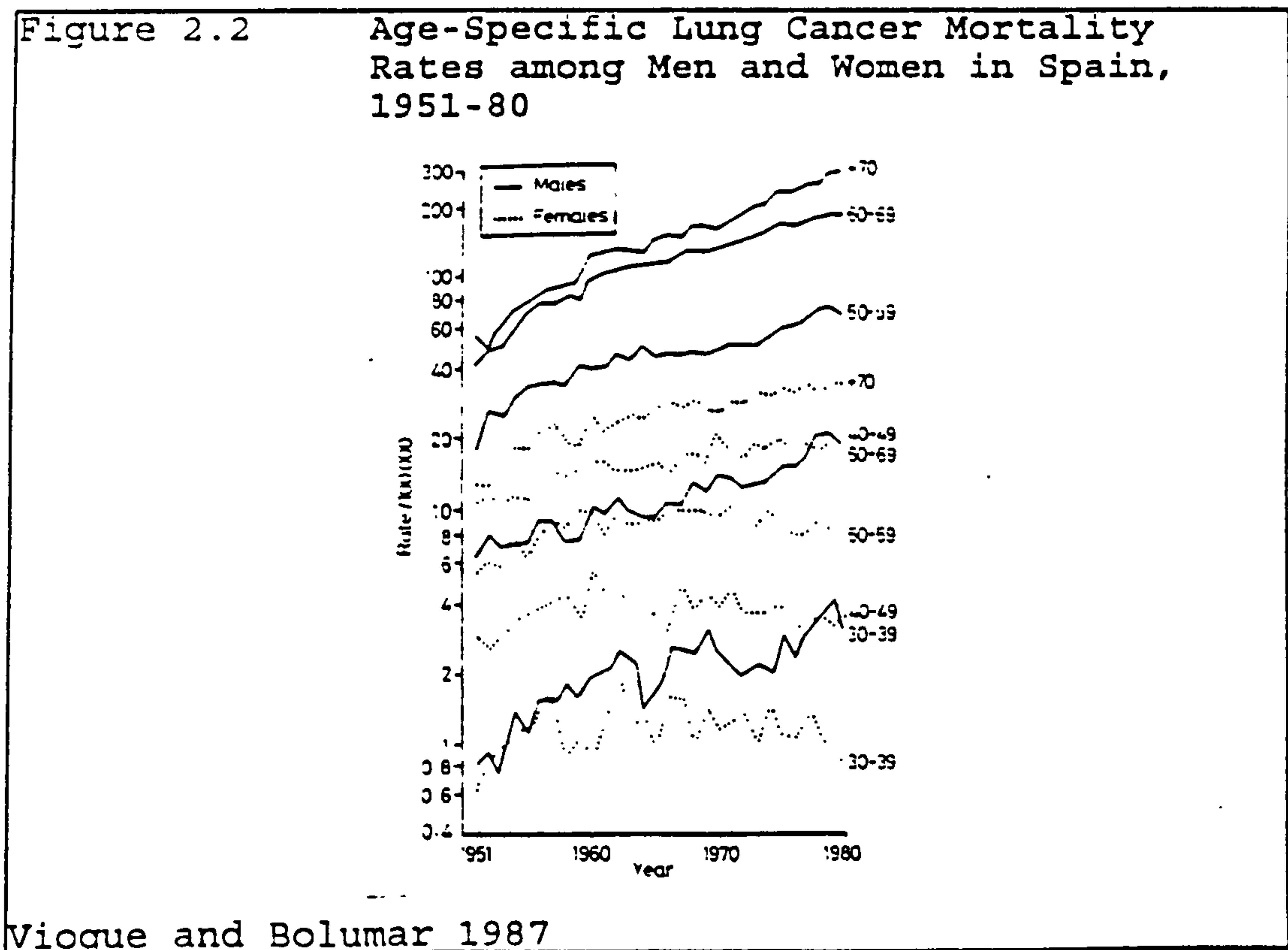
#### 2.2.2.2 Age-Specific Rates



Trends in cancer incidence over calendar period may differ among age groups. For example in Ontario, Canada lung cancer incidence amongst males aged 15-44 years decreased between 1969 and 1982, while the incidence rates for males aged 45-64 years increased over the same



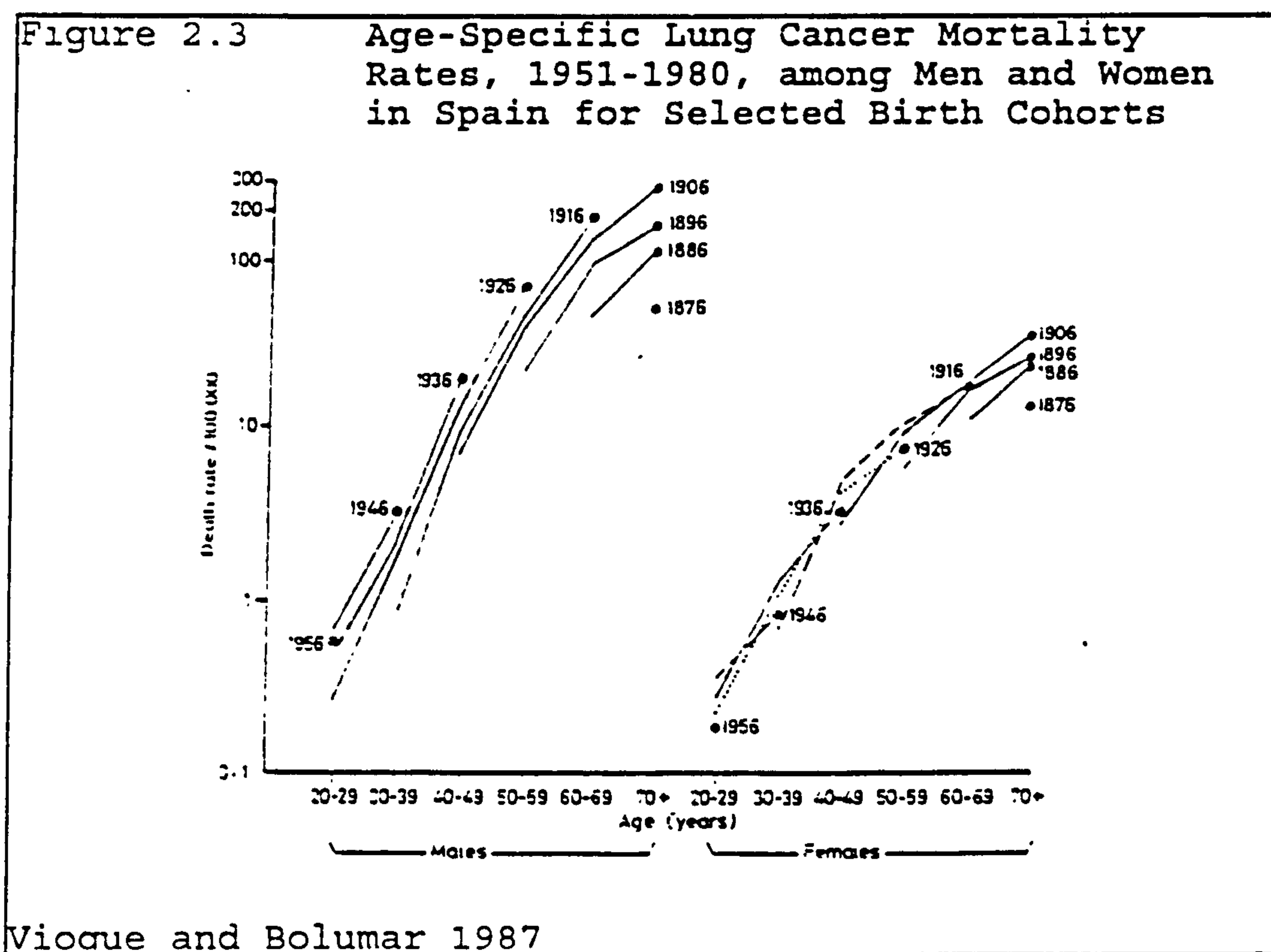
period (Marrett et al 1986). In order to determine the number of incident cancer cases in Ontario for the years 1984 to 1994, the log of the age-specific incidence rates for the age groups 0-14 years, 15-44 years, 45-64 years and 65+ years are regressed against calendar year. Data on cancer incidence for the years 1969 to 1982 are used in the regression (Figure 2.1). The projected rates are applied to the population projections for the period 1984 to 1994, from Statistics Canada. Comparison with newly published data for the years 1983 and 1984 show good agreement between predicted and observed rates.



Trends in malignant melanoma mortality rates in Sweden (Thorn et al 1992) and lung cancer mortality in Spain (Vioque and Bolumar 1987) have been analysed using 10

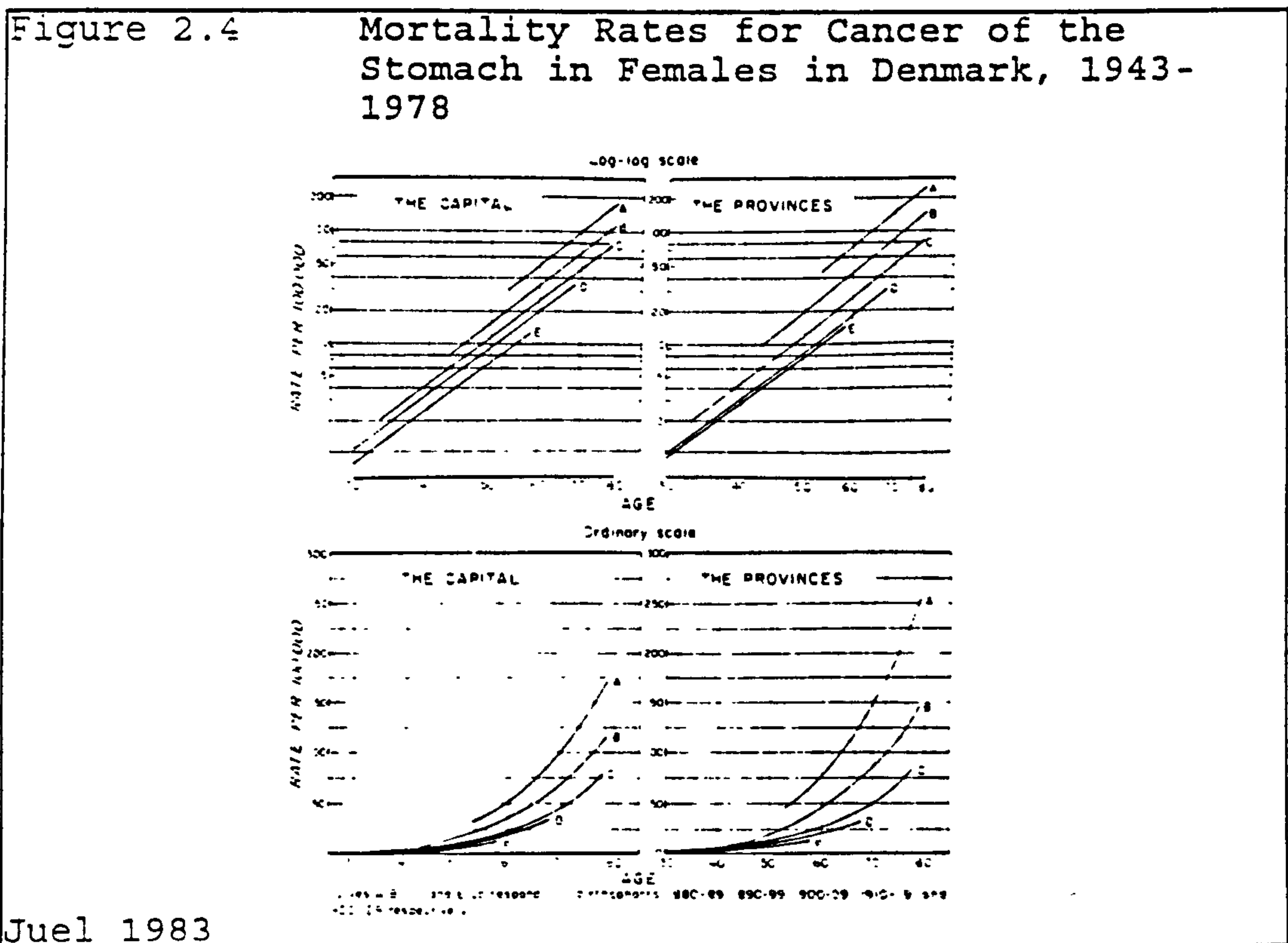


year age-specific rates. The mortality rates in both these studies show that the pattern of increase over calendar period vary with age (Figure 2.2). For lung cancer in Spain males of all ages and females age 60 years and over display upward trends for the years 1951 to 1980. Females aged less than 60 show a increase in mortality rates for the period 1951 to 1960. After 1960 the rates level off and even begin to decrease.



Vioque and Bolumar (1987) also investigate changes in the age-specific lung cancer mortality rates over birth cohort for cohorts born between 1876 and 1956. For males the age-specific rates are increasing over all the birth cohorts except the most recent cohort. However, the only age-specific mortality rate available for this most

recent cohort is for the 20-29 year age group, where the number of deaths would be small (Figure 2.3). Hence the estimate for this cohort may be subject to large errors. For the females, the mortality rates increase over the cohorts born before 1916, and then begin to decrease slightly. Juel (1983) fitted mortality rates for fixed sets of sex, cohort and residence for several cancer sites, using a linear log-log relationship with age (Figure 2.4).



For those sites, such as prostate and stomach, where such lines are parallel, it is possible to determine the trends over the cohorts, as well as to investigate residential differences. However, for lung cancer, and larynx cancer, the age-specific curves are not parallel, the more recent cohorts have a steeper slope and

increasing mortality. For breast, uterus and ovarian cancer it is not possible to fit straight line to the age-specific data because of the presence of a hook in the curve around the time of menopause (Clemmesen 1965).

Projections based on the extrapolation of trends in age-specific rates assume a constant linear trends over the period of interest. If the trends in the rates on which the extrapolation is to be based are not linear, a transformation has to be found which linearises the trends in age-specific rates over time. The choice of this transformation may have significant implication for the projections. This is especially true if the period of projection is long compared to the period for which there is data. While these methods allow age-specific rates to be determined the estimate of the total number of cancer cases obtained may not coincide with those obtained from such measures as the age-standardised or crude rates. This could have serious implication if the total number of cancer cases were required for such purposes as planning. Another problem with this method is that for certain cancers the numbers in each age group may be small, and thus the regression equation is estimated with poor accuracy.

### 2.2.2.3 Age-Period-Cohort Modelling

A simple extension of the age-specific trend analysis is to model either age and period effects or age and cohort effects, assuming additivity of the log scale (Clayton and Schiffers 1987a).

$$\text{Age-Period Model } \ln(R_{ij}) = \alpha_i + \beta_j$$

$$\text{Age-Cohort Model } \ln(R_{ij}) = \alpha_i + \gamma_k$$

where

$i$  is the indicator for age group  $i$ ,  $i=1, \dots, I$

$j$  is the indicator for calendar period  $j$ ,  $j=1, \dots, J$

$k$  is the indicator for cohort  $k$ ,  $k=I-i+j$ ,  $k=1, \dots, K$

$R_{ij}$  is the expected age-specific rate for age group  $i$  and calendar period  $j$

$\alpha_i$  is the age parameter for age group  $i$

$\beta_j$  is the period parameter for calendar period  $j$

$\gamma_k$  is the cohort parameter for cohort

In the phraseology of Case (1956) the age parameters reflect that component of cancer incidence "...which is biological and inescapable.."; the period parameters reflect "...that which is due to constantly changing environment.." and the cohort parameters reflect "...that which is due to early nurture..".

If age and period effects are modelled the age-specific curves should be parallel for the calendar time periods. Similarly if age and cohort effects are modelled, then



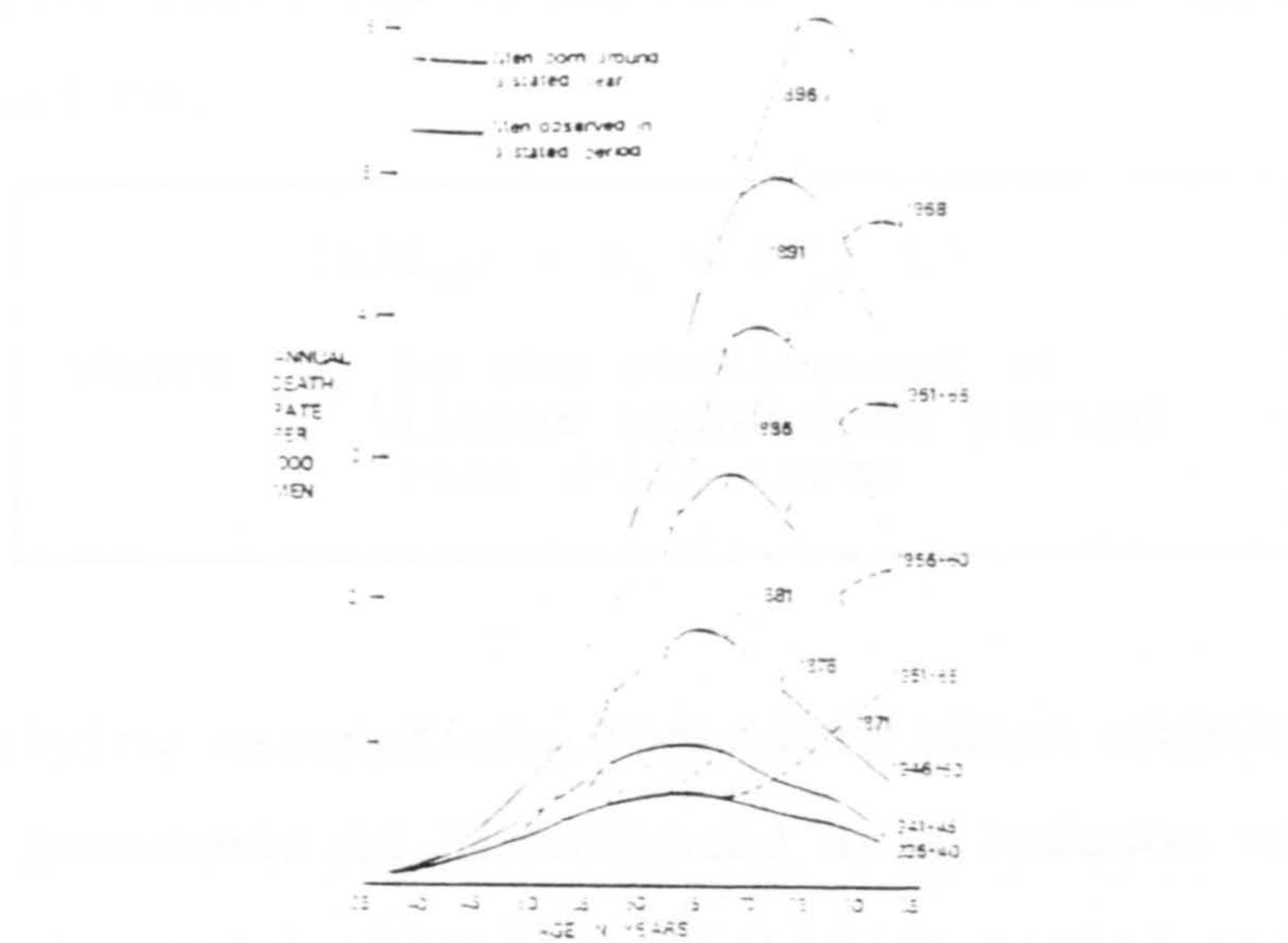
the age-specific curves should be parallel for the birth cohorts. For these models only the first order differences of the parameters are estimable (Section 8.1). These first order differences can be considered to be the log of the relative risks between adjacent ages, period or cohorts (Clayton and Schiffers 1987a). Age-period and age-cohort models do not constrain the relationship between the age-specific rates and the time variable to be linear over the whole time period under consideration.

For some cancers there may be rational arguments for there to be both period and cohort effects in the model (eg male lung cancer in England and Wales for the period 1935-68, Doll 1971), as well as age effects (Figure 2.5). For these cancer sites an appropriate model would be:

$$\ln(R_{ij}) = \alpha_i + \beta_j + \gamma_k$$

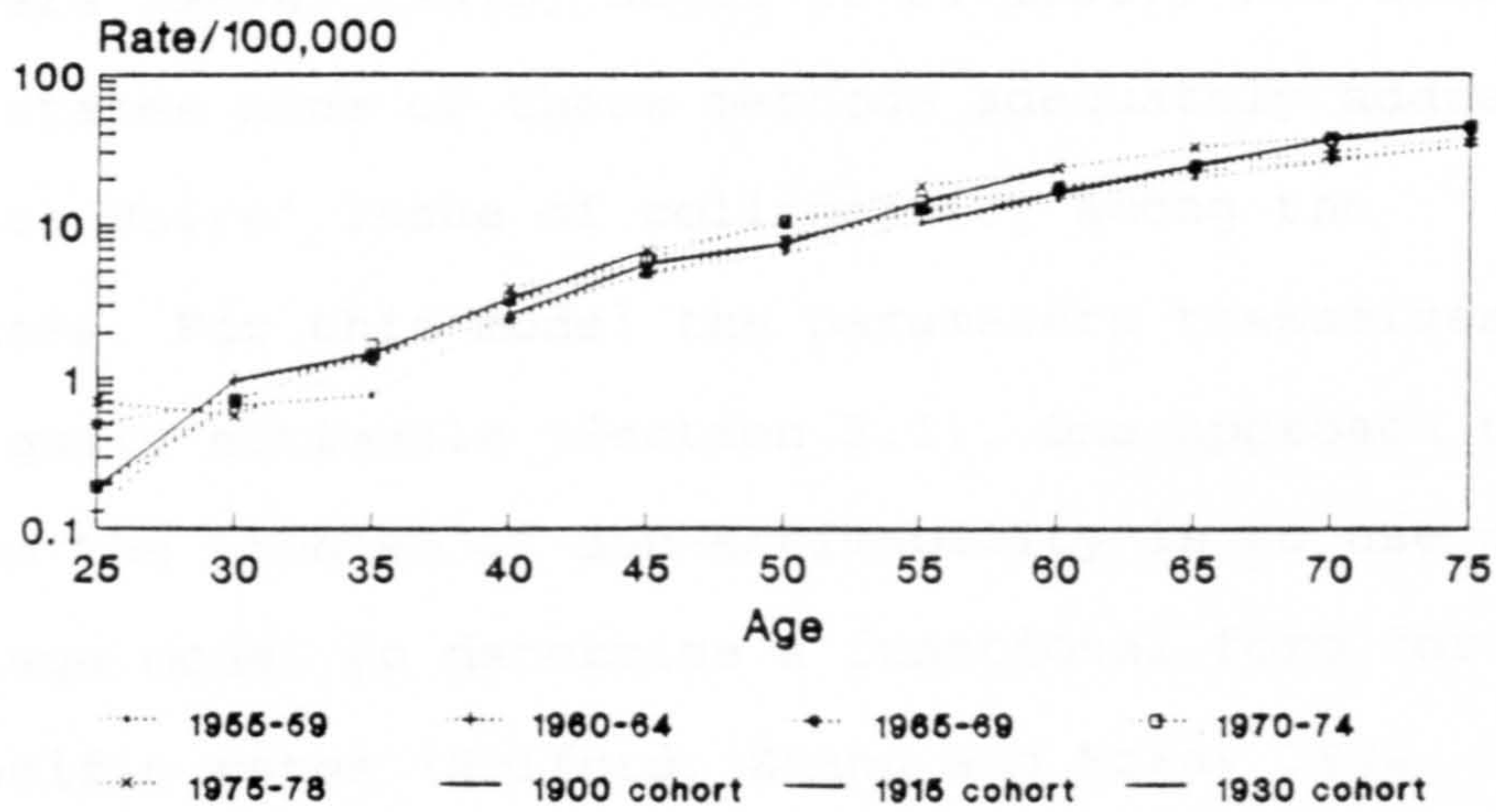
For other cancer sites, there may be no clear cut indication as to whether the observed trends are due to

Figure 2.5 Lung Cancer Mortality among Males in England and Wales, 1935-1968, by Time Period and Year of Birth. From Doll (1971)



Day and Charnay (1982)

Figure 2.6 Age-Specific Lung Cancer Mortality Rates in Belgium, 1955-1978 Females



Source of data: WHO Mortality Database  
Clayton and Schiffers (1987a)

period and/or cohort effects (eg female lung cancer in Belgium during the period 1955-78 Clayton and Schiffllers 1987a Figure 2.6). For these cancer sites an appropriate model would be:

$$\ln(R_{ij}) = \alpha_i + \delta_j(j-j_0)$$

where  $\delta_j$  is the coefficient of linear trend over period (the drift term)

The underlying assumptions regarding these models are that the incidence is distributed as a Poisson variable and that the model relating age and/or period and/or cohort is additive on the log scale. Various methods have been suggested to estimate the age, period and cohort parameters (Barrett, 1978, James and Segal 1982, Osmond and Gardner, 1982; Holford 1983, 1985, Clayton and Schiffllers 1987a, 1987b, Negri et al 1990), but as Kupper (1984) states none of these methods adequately address the 'Bete Noire' issue of collinearity among the parameters. For this model the parameters themselves are not uniquely estimable (Section 8.1). One approach to overcome the problem of non-estimability is to use of the multistage model to determine a functional form for the age-specific rates (Holford, Zhang and McKay 1994, Morrison et al 1995). While this approach overcomes the problem of non-estimability there is concern that choice of the wrong functional form for the age-specific rates may lead to incorrect interpretations of the trends over

calendar period and birth cohort.

If standard age-period-cohort analysis are used the estimable functions of the parameters are the second order differences (Rodgers 1982). Investigations of these second order differences over the three dimensions of time may help to understand the trends in cancer incidence or mortality (Holford, 1983). Despite these problems, age-period-cohort analysis have been used extensively to describe trends in cancer incidence, the method used being dictated somewhat by geographical proximity to the originator of the idea (Osmond, Gardner and Acheson 1983, Rousch et al 1985).

Clayton and Schiffers (1987a,1987b) in review articles, discuss the various approaches to age period cohort analysis. They suggest a hierarchy of models to determine whether the rates depend on period and/or cohort (Table 2.1).

If the relationships of incidence to calendar period and birth cohort are essentially linear in the log scale, the trends can be modelled using a linear relationship with either period or cohort by the inclusion of a drift term ( $\delta$ ) (Model 2, Section 8.1). If this is the case then it is not possible to identify whether the time trends are due to period or cohort effects. If non linear period and



Table 2.1 Hierarchy of Model Fitting

Model 1	Age	
	$\alpha_i$	
	↓	
Model 2	Age + Drift	
	$\alpha_i + \delta_j(j-j_0)$	
	↙ ↘	
Model 3a	Age + Drift + Period	Model 3b Age + Drift + Cohort
	$\alpha_i + \delta_j(j-j_0) + \beta_j$	$\alpha_i + \delta_j(j-j_0) + \gamma_k$
	↘ ↙	↙ ↘
Model 4	Age + Period + Cohort	
	$\alpha_i + \beta_j + \gamma_k$	

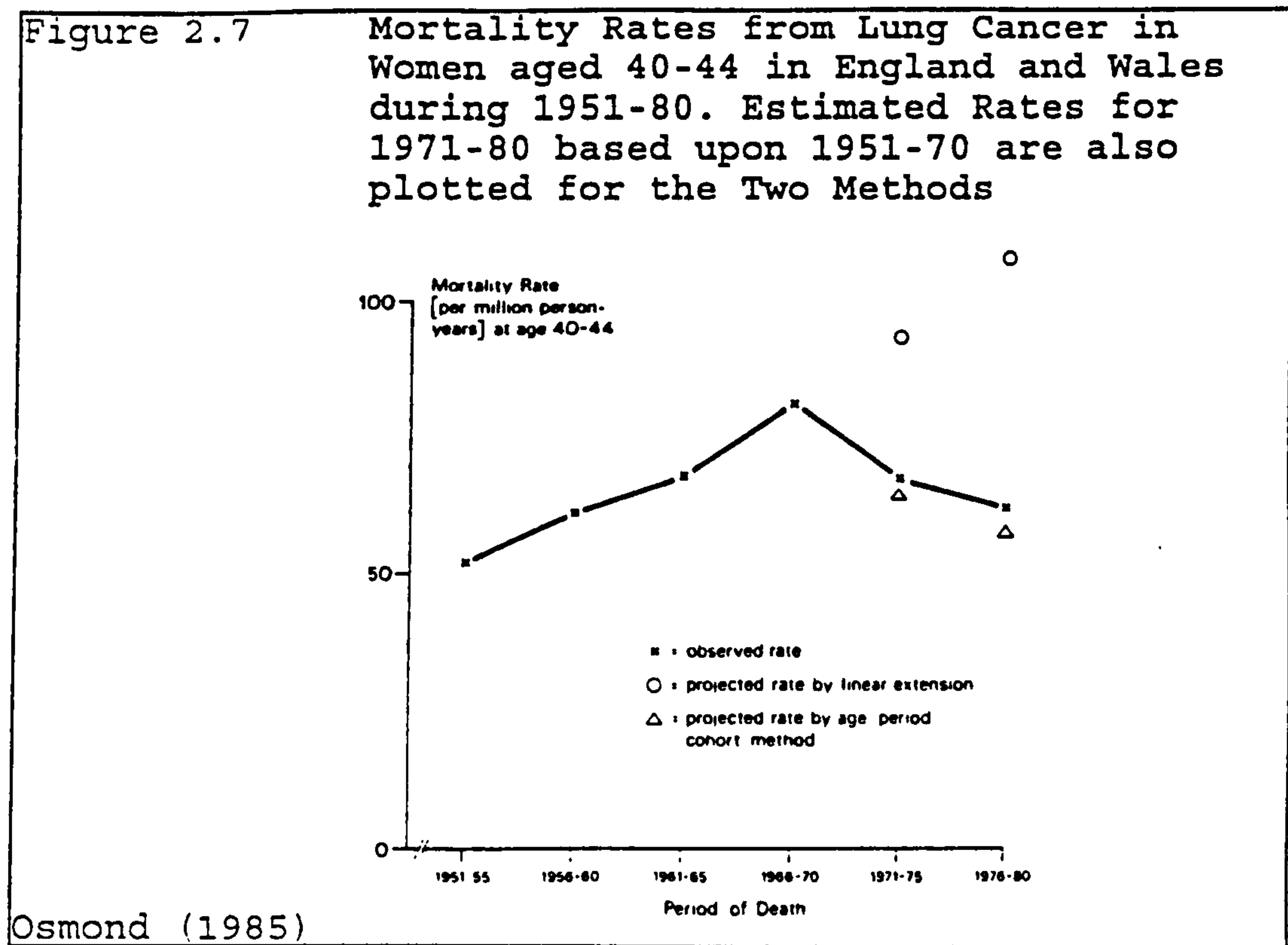
cohort effects contribute to the fit of the model (Model 4), then the second order differences of the parameter estimates, which measure the curvature of the parameter curve, are the comparative log relative risks among periods and/or cohorts (Section 8.1). The change in the scaled deviance with the addition of each set of parameters, can be used to determine the contribution of the age, period and/or cohort parameter sets (Section 8.1). The change in scaled deviance is approximately distributed as a chi square, with degrees of freedom equal to the number of parameters in the set being tested. However, these tests of significance may not be particularly sensitive to smoothly increasing or decreasing trends over time (Clayton and Schifflers 1987a).

In fitting any statistical model to data it is always important to ensure that the fit of the model is adequate (McCullagh and Nelder 1989). The size and pattern of

residuals indicate how well the model fits the data. If the residuals show no systematic pattern and no extreme values, then it can be assumed that the fit is adequate. If the residuals are randomly distributed but have greater variation than would be expected, then it is possible that the data arise from clustered Poisson process (Breslow 1984). The model would still be appropriate, but the standard errors of the parameter estimates have to be multiplied by the overdispersion parameter. This is calculated as the square root of the final scaled deviance divided by the degrees of freedom (McCullagh and Nelder 1989).

The parameter estimates from age-period-cohort analysis have been used to predict cancer incidence (Osmond 1985, Negri et al 1990). Osmond suggests fitting an age-period-cohort model to mortality data and obtaining future period and cohort parameter estimates by linear extrapolation of the trends over the more recent period and cohort parameters estimates. The decision as to the number of past values to include in the regression and the form of the regression is arbitrary. The estimated age parameters, and the extrapolated period and cohort parameter values are used to estimate the predicted incidence. In England and Wales the projections for female lung cancer mortality in the age group 40-44 years were more accurate if the trends over both calendar period and birth cohort were included in the projection

model (Figure 2.7, Osmond 1985).



Negri et al (1990) also use age-period-cohort models for predicting future mortality due to cancer. The projections are based on age-specific mortality data for the years 1951 to 1984. The age parameters are assumed to remain constant over the projection period. Due to the low numbers of deaths in the younger age groups and hence recent cohorts, they constrain the future cohort parameters to remain equal to that for the last cohort (1955). Three scenarios are used for the period parameters; the period parameters remain equal to that for the last period (1980-84), increase linearly determined by the last seven period estimates, or reflect recent trends in risk behaviour, or changes in treatment. They observe that if the period parameter estimates



reflect the smoking trends, then the projected mortality for lung cancer is less among males, and higher among females than if extrapolations of trends in the parameter estimates are used in the projections.

Thorn et al (1992) show that if the age-period-cohort model is used to project mortality of malignant melanoma in Sweden, the rate of increase of the estimated age-standardised rates would decline over the 20 year period following 1987. This follows the pattern shown by the age-standardised rates, except that the increase is more marked in the results from the age-period-cohort analysis. They also investigated the effect of including differing numbers of period and cohort parameter estimates in the projection equations for these estimates. For malignant melanoma, the number of parameter estimates included in the regression has little effect on the estimated age-specific mortality rates. The only exception to this is the period parameters for females. In this instance the trends in malignant melanoma for females is not linear over the recent years, with a steep increase in mortality in the late 1970s.

The methods described above all require the number of incident cases or deaths for each sex, age and year of diagnosis or death category. Age-specific population figures are also required for each sex. In order to extrapolate the trends in rates, the rates may have to be



transformed to obtain a linear relationship with time. The choice of this transformation may have significant implications on the projected rates. If age-standardised rates are used for projections it is assumed that the trends over calendar period are similar for each age group. It is also assumed that the changes in rates are due to changes over calendar period and not birth cohort.

Age-period-cohort analysis allows the investigation of the contribution of period and/or cohort effects to the model. However, if the age-period-cohort model is used the parameter estimates are not unique, and must be interpreted with care. In order to estimate future rates using age-period-cohort analysis, future period and cohort parameters have to be estimated. This may be achieved by linear extrapolation of the parameter curves. A problem will arise if these curves are non linear for the recent periods or cohorts. This problem will be more marked with period effects than cohort effects (Thorn et al 1992). The incidence or mortality rates are lower in the younger age groups for which the extrapolated cohort parameters are estimated. The above models also assume that past trends in the rates will continue for the period of projection. Negri et al (1990) attempt to investigate the effects of changes in trends by constraints placed on future period parameter estimates reflecting smoking behaviour.

Hakulinen et al (1986) compare the projections calculated by Teppo et al (1974) to the actual number of cancer cases diagnosed in 1980. For 8 out of the 12 cancer sites investigated the incidence in 1980 does not fall within the 90% confidence intervals for the projected cancer incidence; for example lung cancer incidence is over estimated. They suggest that the inaccuracy of the projections were due in part to changes in risk behaviours, such as smoking behaviour, introduction of screening programs (breast and cervical cancer), and changes in diagnostic methods or criteria (cancers of the urinary tract). To overcome the problems due to changes in risk behaviours, models for lung cancer incidence and/or mortality have been developed incorporating the trends in smoking behaviour (Hakama, Hakulinen and Laara 1986, Stevens and Moolgavkar 1979 1984, Brown and Kessler 1988).

### **2.2.3 Inclusion of Risk Factor Information in the Models**

This section will discuss the inclusion of risk factor information in the models used to project future cancer incidence. Hakama et al (1986) state that 'If the latent period between first exposure and the diagnosis of cancer is long, trend data on major risk factors can sometimes be used to predict incidence.' Therefore models for trends in incidence or mortality rates, that incorporate

trends in risk behaviour and the latent, or lag, period between exposure and incidence are discussed.

It has been observed that the trends in lung cancer incidence and/or mortality are consistent with trends in smoking habits (Vioque and Bolumar 1987, La Vecchia et al 1988, Hakama et al 1989). In Spain the trends of the age-specific lung cancer mortality over cohorts are similar to the trends of both the percentage of the population who smoke and the consumption of cigarettes by males (Vioque and Bolumar 1987). Similarly in Switzerland the trends in lung cancer mortality are similar to the patterns of smoking behaviour measured by the percentage of the population who smoke, the distribution of smokers by age, sex and number of cigarettes smoked per day, and the distribution of smokers by age sex and the age at starting to smoke (La Vecchia et al 1988). The average tar content of cigarettes is also important in the interpretation of trends in lung cancer (La Vecchia et al 1988). In Finland trends in lung cancer mortality are similar to cohort specific trends in smoking, particularly if the introduction of low tar and filter cigarettes are accounted for (Hakama et al 1989). These authors note the relationship between the various smoking behaviours and lung cancer mortality, and use these relationships to help interpret the trends. However, they make no attempt to incorporate the information on smoking behaviour into models to explain the trends in lung



cancer mortality rates. There is also no discussion of the lag times between smoking behaviour and mortality from lung cancer.

A simple method to incorporate the trends in smoking behaviour is to constrain projected cohort or period parameter estimates to reflect changes in smoking behaviour (Negri et al 1990). These constraints can also incorporate an estimated lag period, so that changes in smoking behaviour affect lung cancer mortality or incidence after a given number of years. In the example used by Negri, the period parameter estimates were constrained to decrease for males and increase for females, reflecting the recent patterns in smoking behaviour.

Hakama and Pukkala (1984) determine the lag times between the age-standardised incidence of lung cancer and smoking, by correlating the second order differentials of the incidence of lung cancer and the number of cigarettes smoked per day per adult (Harrison and Stevens 1976). The correlations were calculated for different lag times. Predictions were then based past trends in smoking, using those lag times for which the correlations were significant (9 years, 10 years, 21 years and 23 years).

LIVERPOOL  
UNIVERSITY  
LIBRARY





The resulting estimating equation is:

$$R^s_j = 24.2 + 17.6x_{j-9} + 10.6x_{j-10} + 10.3x_{j-21} - 14.3x_{j-23}$$

where

$R^s_j$  is the expected age-standardised incidence rate in year  $j$

$x_j$  is the number of cigarettes smoked per adult per year in year  $j$

This model is based on the average cigarette consumption for the population and therefore takes no account of the percentage of smokers who stop smoking, or the percentage of non-smokers who take up smoking. The authors suggest that since the effect of stopping smoking is relatively rapid, that the estimates from this equation would tend to be too high. They also warn that correlation analysis may not be sensitive enough to determine the lag period when the trends in cancer incidence are smooth. There may also be difficulties with the quality of data for both lung cancer and smoking, so that estimates of second differentials may be subject to extraneous variation. This model assumes that smoking behaviour affects all age groups in the same manner. It is generally believed that an individual's smoking habits are acquired early in life (Hammond 1966) and vary with birth cohort (Doll and Hill 1964). The percentage of the population who smoke also varies with birth cohort (Harris 1983). Thus changes in smoking behaviour are more likely to relate to cohort effects than period effects.

Stevens and Moolgavkar (1979, 1984) fitted a series of models that related lung cancer mortality to age, period, cohort, sex and smoking effects. They assumed that the number of deaths are Poisson variates. The set of models they investigated were (Stevens and Moolgavkar 1979):

**Model 1:**  $D_{iks} = A_i B_j C_k N_{iks}$

**Model 2:**  $D_{iks} = A_i B_j C_k S_s N_{iks}$

**Model 3:**  $D_{iks} = A_i B_j C_k [1 - P_{iks} + P_{iks} X^{G_{iks}}]$

where  $D_{iks}$  is the expected number of deaths in age group  $i$ , cohort  $k$  and sex  $s$

$A_i$  is the effect of age group  $i$ ,

$B_j$  is the effect of period  $j$ ,

$C_k$  is the effect of cohort  $k$

$N_{iks}$  is the population at risk in age group  $i$ , cohort  $k$  in sex  $s$

$S_s$  is the effect of sex  $s$

$P_{iks}$  is the proportion of the population  $N_{iks}$  that has ever smoked,

$X$  is the relative risk associated with smoking 1 unit (= 20 cigarettes per day for 20 years), and

$G_{iks}$  is the average cumulative consumption of cigarettes per smoker (in units) in age group  $i$ , cohort  $k$  of sex  $s$ .

Model 1 assumes that there is no difference between males and females, and that all the variation may be explained by age, period and cohort effects. Given that the age-specific lung cancer mortality rates are substantially lower in females than males for the period of interest, it is no surprise that this model does not fit the data

adequately. Inclusion for a factor for sexual differences improved the fit markedly. The effect for males is seven times that for females. This difference was well explained by the differences in smoking history (Model 3). An advantage of this method is that it allows the estimation of the relative risk of dying from lung cancer due to smoking 20 cigarettes per day for 20 years (1 unit). The estimate that Stevens and Moolgavkar obtain (4.2) is similar to those obtained from other studies (Doll and Hill 1964, Hammond 1966). Model 3 also allows estimation of the mortality rates for lung cancer among non smokers from the quantity  $(A_i B_j C_k [1 - P_{iks}])$ . Lung cancer mortality among non smokers is estimated to peak about the time of the Clean Air Act (1956), which may indicate a relationship between air pollution and lung cancer mortality (Section 2.4.3).

The data required for these analyses are the age-specific percentages of the population who smoke for each sex separately, and the total cigarette consumption for each sex and age group by birth cohort. The percentage of the population smoking is available from 1948 for each sex by broad age groups (Section 3.4). Data for 5 year age groups must be estimated. The cumulative consumption data are adjusted for the decrease in the average tar content of cigarettes, but do not take account of differential mortality between smokers and non-smokers (Stevens and Moolgavkar 1984). In the model there is no consideration



of the lag period between consumption and onset of disease. If this lag period is of the order of 20 years then the cumulative consumption figures used in this analysis will over estimate the smoking effect more in the younger age groups than in the older age groups. This may explain the difference in relative risks obtained when the model is fitted separately to the age ranges 35-59 years and 60-84 years. The relative risk for the younger age group is 3.04 and that for the older age group is 4.2. The model is based on average smoking behaviour within the cohorts. In more recent years there have been more anti smoking campaigns, a change towards filter cigarettes and a decrease in the tar content, which may lead to wide variation in smoking behaviour within a cohort. With this model it is not possible to evaluate these changes in smoking behaviour.

Brown and Kessler (1988) also incorporated an index of cigarette consumption in the age-period-cohort model. The model they propose replaces the period parameters with a measure of the changing tar content of cigarettes.



The model is:

$$\log(R_{ij}) = A_i + BX_j + C_{I-i+j}$$

where

- $R_{ij}$  is the expected mortality rate for age  $i$  period  $j$
- $A_i$  is the age effect for age  $i$
- $X_j$  is the measure of the populations exposure to cigarette tar
- $C_{I-i+j}$  is the cohort effect for cohort  $I-i+j$

Two measures of exposure to cigarette tar were investigated :

- the average tar content of cigarettes and
- the product of average tar and average number of cigarettes sold per capita (age>20 yrs)

The average number of cigarettes sold per capita in the years of interest 1954 to 1980 had to be estimated for each sex separately. This is estimated from the prevalence rates of smoking by age and sex, obtained from the 1978-80 HIS household survey conducted by the US National Center for Health Statistics (1984), and the total sales of cigarettes in each year. The estimated prevalence rates of smoking were adjusted for the effects of differing mortality among smokers and non-smokers. The lag time between tar consumption and lung cancer mortality is estimated by minimising the final scaled deviance in the model (Section 7.1). The minimum is found

for a lag period of 21 years.

The model gave good fit of the data and overcame the problem of co-linearity inherent in age-period-cohort analysis. The cohort values closely mimicked the smoking prevalence data for the individual birth cohorts. The slope of the relationship with the measure of the population's exposure to tar was larger in females than males. When the analyses are repeated separately for those aged less than 50 years and those aged 50 years or older, the slopes are larger in the young age group than in the older age group. The rationale for these results is that males and older people tend to continue to smoke higher tar cigarettes (National Center for Health Statistics 1984). Thus decreases in tar would have a larger effect on females and young people. The final model that they used for projections of lung cancer mortality incorporated the differing relationships for both age and sex. For each sex the model is:

$$\log(R_{ij}) = \begin{cases} A_i + B_1 X_j + C_{I-i+j} \\ A_i + B_2 X_j + C_{I-i+j} \end{cases}$$

where

$B_1$  is the regression coefficient  
for ages 30-49 years and

$B_2$  is the regression coefficient  
for ages >50 years

The differences in the relationships between tar

consumption and lung cancer incidence may indicate that average tar consumption should not be considered only a period effect. Smoking patterns are thought to be acquired for life, and thus the type of cigarettes smoked does not necessarily change due to an overall change in tar content of cigarettes. Therefore it may be more sensible to incorporate a measure of cigarette consumption in the cohort parameters, and only use the average tar content of cigarettes in the period parameters.

In order to project future mortality rates using the above model, estimates of the average tar content of cigarettes, the number of cigarettes sold to each sex in the United States and the percentage of young adults who smoke are required. The relationship between the cohort parameter estimates and the prevalence of smoking among young adults is used to estimate future cohort parameters. This relationship allows investigation of the effects of changes in smoking prevalence on mortality from lung cancer. In particular, Brown and Kessler (1988) investigate the effects of attainment of the National Cancer Institute Year 2000 Project objectives, to reduce smoking prevalence to 15% in young adults by the year 2000 (Greenwald and Sondik 1986). They also investigate differing projections for future average tar content of cigarettes, linear decline or constant 1982 values. The projected mortality rates are not affected by differing



assumptions related to the smoking variables in the near future. However the differences become apparent as the more recent cohort age.

In all these analyses the risk behaviour in question may not be a causative risk factor, but may be an indicator for trends in such risk factors. Using risk behaviour data instead of actual trends in causative factors, to improve the accuracy of projections, should not affect the projections as long as the relationship between indicator condition and causative risk factor remains stable.

If risk factors are to be included in the projection models for lung cancer, it is necessary to determine the factors that have the highest association with the development of lung cancer. The next section briefly reviews the literature on the aetiology of lung cancer. As has been discussed any model incorporating risk factor information, requires good historic data on the chosen risk factors. The availability of such information for the projections for Mersey Region will be discussed in Chapter 5.

### **2.3 Aetiology of Lung Cancer**

In this section the aetiology of cancer of the male and female lung will be discussed. The most important risk



factor is known to be cigarette smoking, accounting for 85% of lung cancer deaths in the developed world (Peto et al 1992). Other potential risk factors are occupational exposure, air pollution, diet and social class (Tomatis 1990). Occupational exposure and air pollution act synergistically with cigarettes smoking (Whittemore and MacMilan 1983, Hornung and Meinhardt 1987), but there are also high correlations between exposure to these risk factors and cigarette smoking (Pastorino et al 1984). The main categories of risk factors are discussed in the following sections.

### 2.3.1 Cigarette Smoking

Cigarette smoking has been considered as a causative agent for lung cancer since the early 1950s (Doll and Hill 1950, Wynder and Graham 1950). Doll and Hill (1950) showed that both male and female smokers had an elevated risk of developing lung cancer. They also demonstrated a dose response relationship with the number of cigarettes smoked per day, the total quantity smoked in a life time and the number of years smoked. There are inverse relationships with the age at starting to smoke and the years since stopping smoking.

For an individual there appears to be a clear cut dose response relationship between number of cigarettes smoked per day and the risk of developing lung cancer. A large

cohort study of British male doctors showed increasing relative risks with increasing quantity smoked per day (Doll and Peto 1976). Similar findings have been reported by Hammond (1966), Garfinkel and Stellman (1988) and Dean et al (1977) (Figures 2.8).

Garfinkel and Stellman (1988) show that the age-adjusted relative risks of dying from lung cancer increase with increasing duration of smoking (Figure 2.9). The relative risks, for a given duration of smoking also increase with increasing daily cigarette consumption, so that for women who have smoked more than 30 cigarettes per day for more than 40 years have a relative risk of 38.8 when compared with nonsmokers of a similar age distribution.

Thus it has been shown that both increasing quantity and duration increase the risk of developing lung cancer. Total consumption may be a better measure of cigarette exposure. Risch et al (1993), in their case-control study, use the pack-years (= number of packs (20 cigs) smoked per day x years of smoking) as a measure of total consumption. After adjusting for years since quitting, odds ratios for the development of lung cancer increase with increasing pack-years, (Figure 2.10). The odds ratios were higher for females than for males.



Figure 2.8a Relative Risk for Lung Cancer according to the Number of Cigarettes Smoked per Day, Male

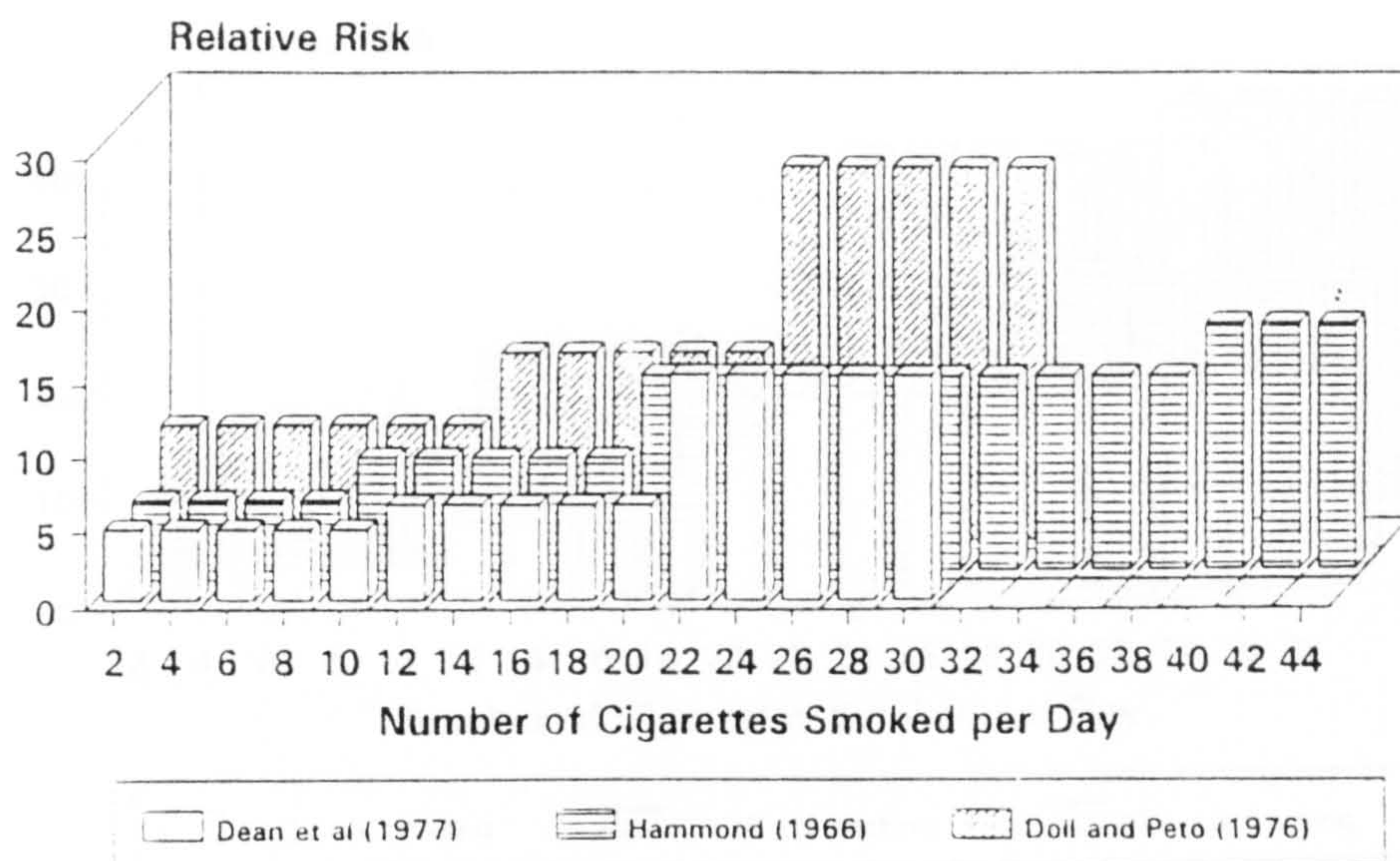


Figure 2.8b Relative Risk for Lung Cancer according to the Number of Cigarettes Smoked per Day, Female

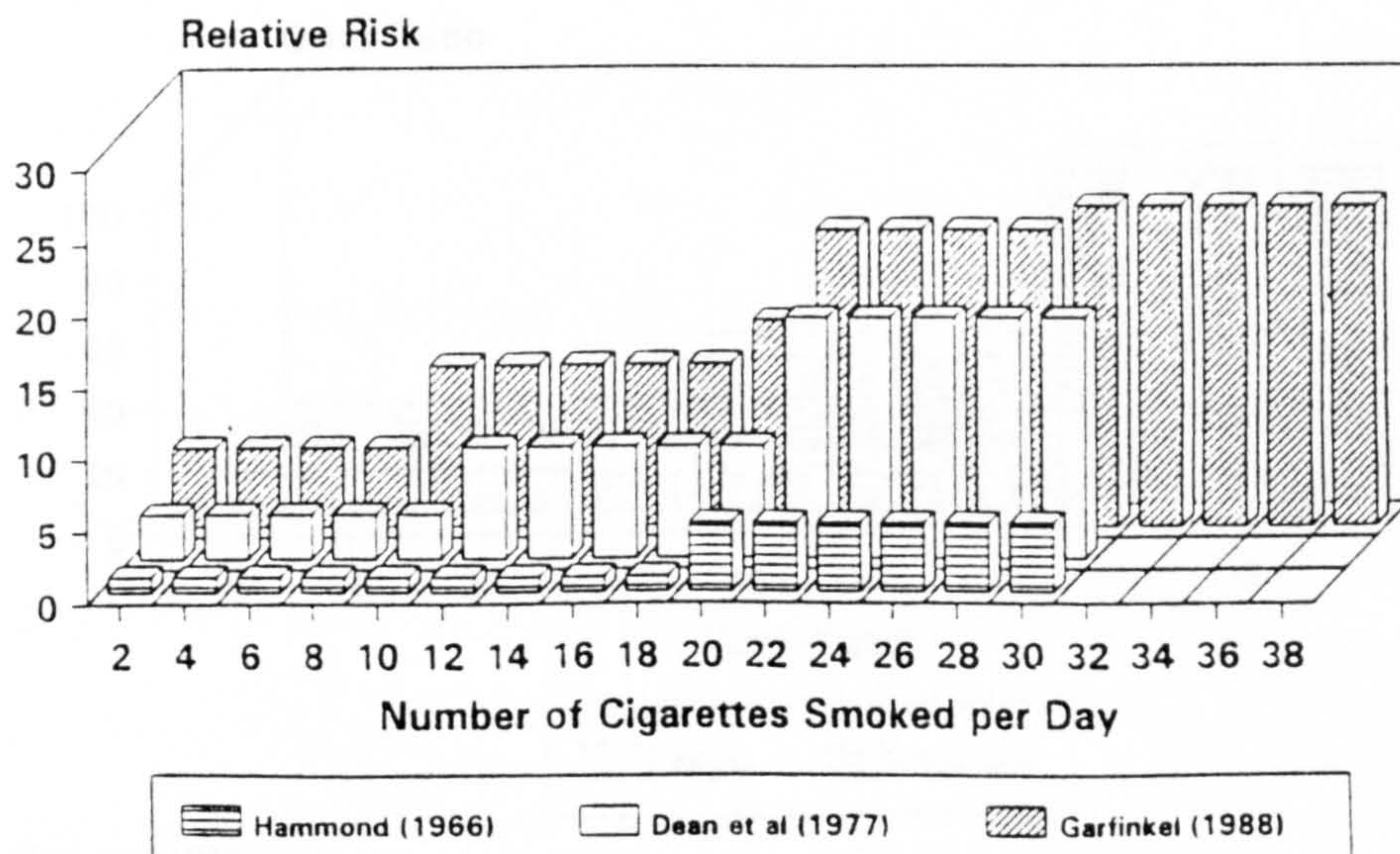
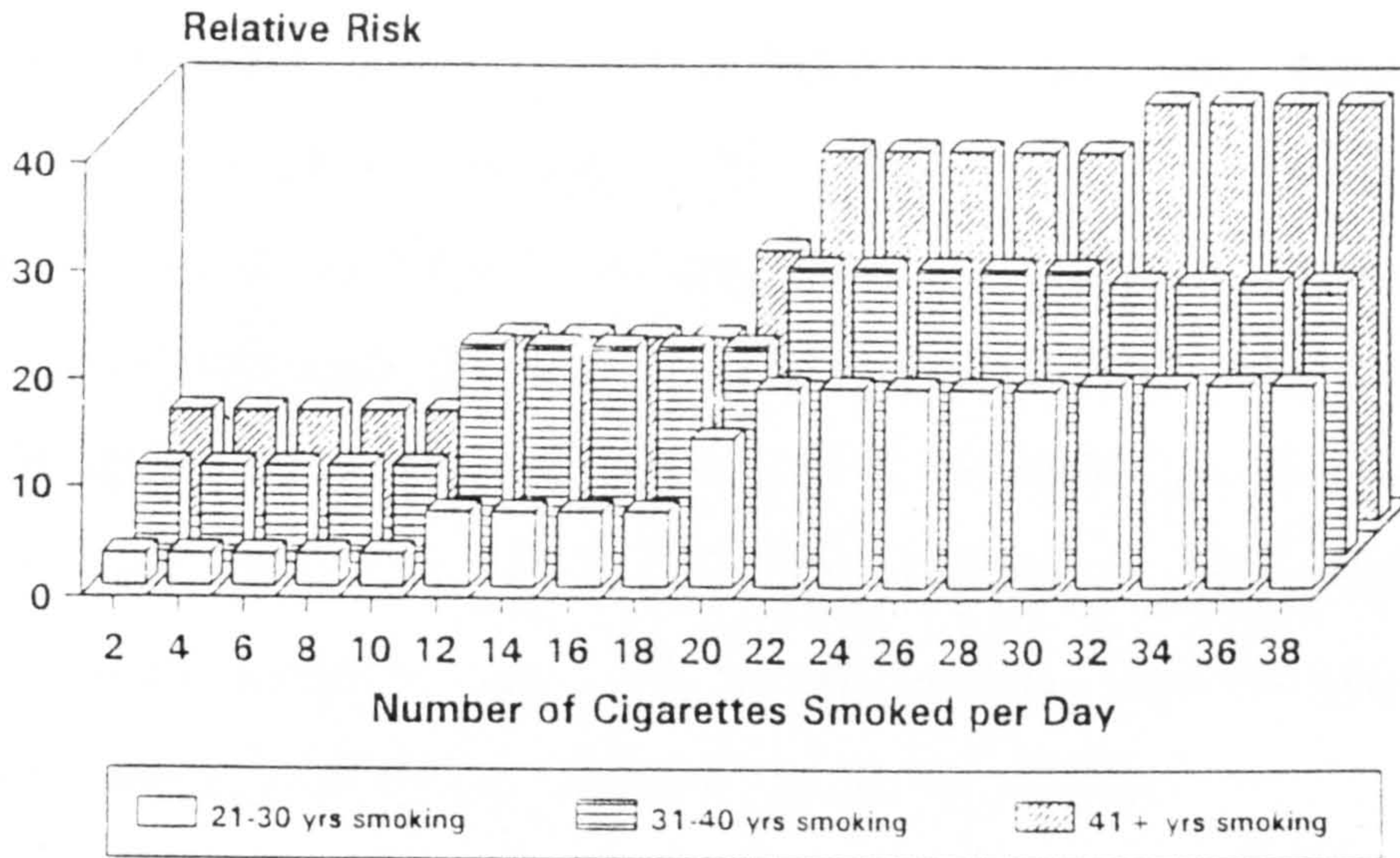


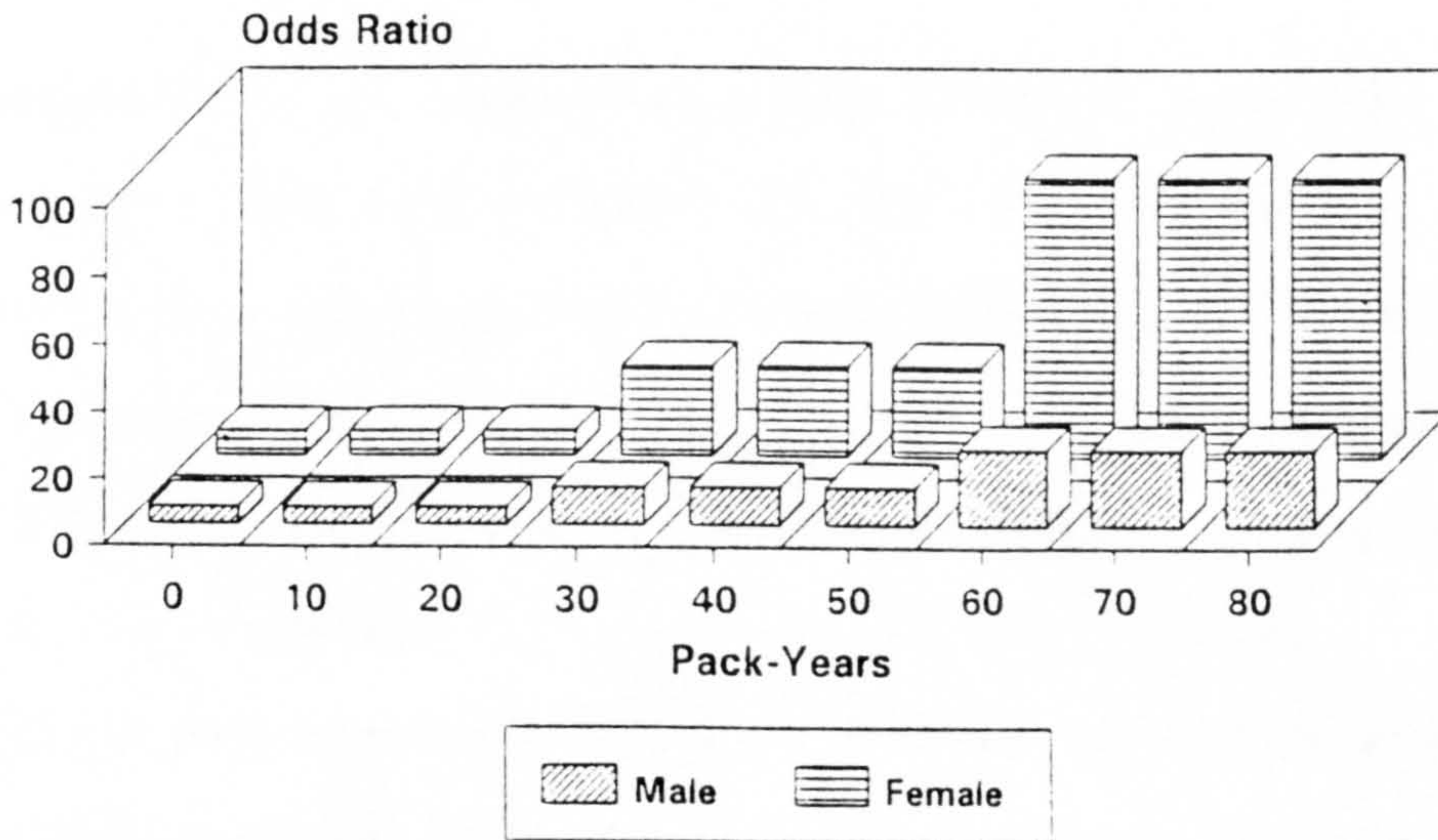


Figure 2.9 Relative Risk for Lung Cancer according to Number of Cigarettes Smoked per Day, and Duration, Male



Garfinkel & Stellman 1988

Figure 2.10 Lung Cancer Risk according to Lifetime Cigarette Consumption Canada 1981-1985



Risch et al 1993  
 Pack-Year = No. Packs(20 cigs)/day x Years  
 Adjusted for Years since Quitting



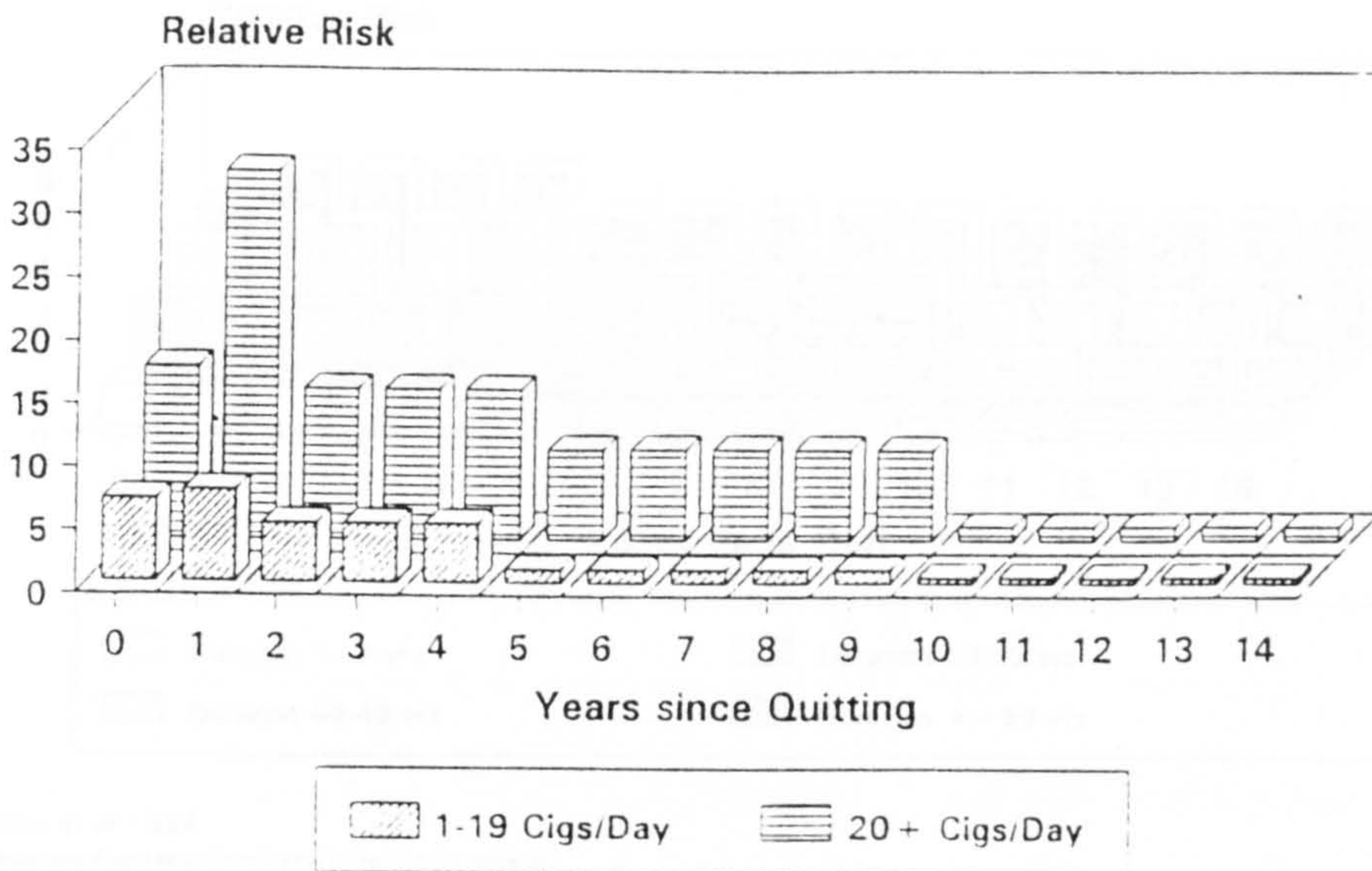
The relative risk for lung cancer decreases with increasing time since quitting smoking (Garfinkel and Stellman 1988, Hammond 1966, Lubin et al 1984a). The decline in risk is slower for those people who have smoked more cigarettes per day (Figure 2.11). Lubin et al (1984a) also show that the decline in risk is slower the longer the history of smoking (Figure 2.12). All the studies show an increase in risk of mortality during the initial period after quitting. This excess is probably due to those people who quit smoking for health reasons and are therefore at a higher risk of dying.

Hammond (1966) show increasing relative risk with decreasing age at which smoking started (Figure 2.13). This may be due to confounding with the duration of cigarette smoking.

The composition of cigarettes has changed over the last two decades. The tar content of the cigarettes has reduced partly through the increasing popularity of filter cigarettes. Lubin et al (1984a, 1984b) showed that the relative risk for any tobacco use was 7.5 for males and 3.9 for females. Lifetime smokers of filter cigarettes had approximately half that relative risk of developing lung cancer. However there is no evidence that switching to low tar cigarettes after a substantial exposure to high tar cigarettes reduces the risk of developing lung cancer.

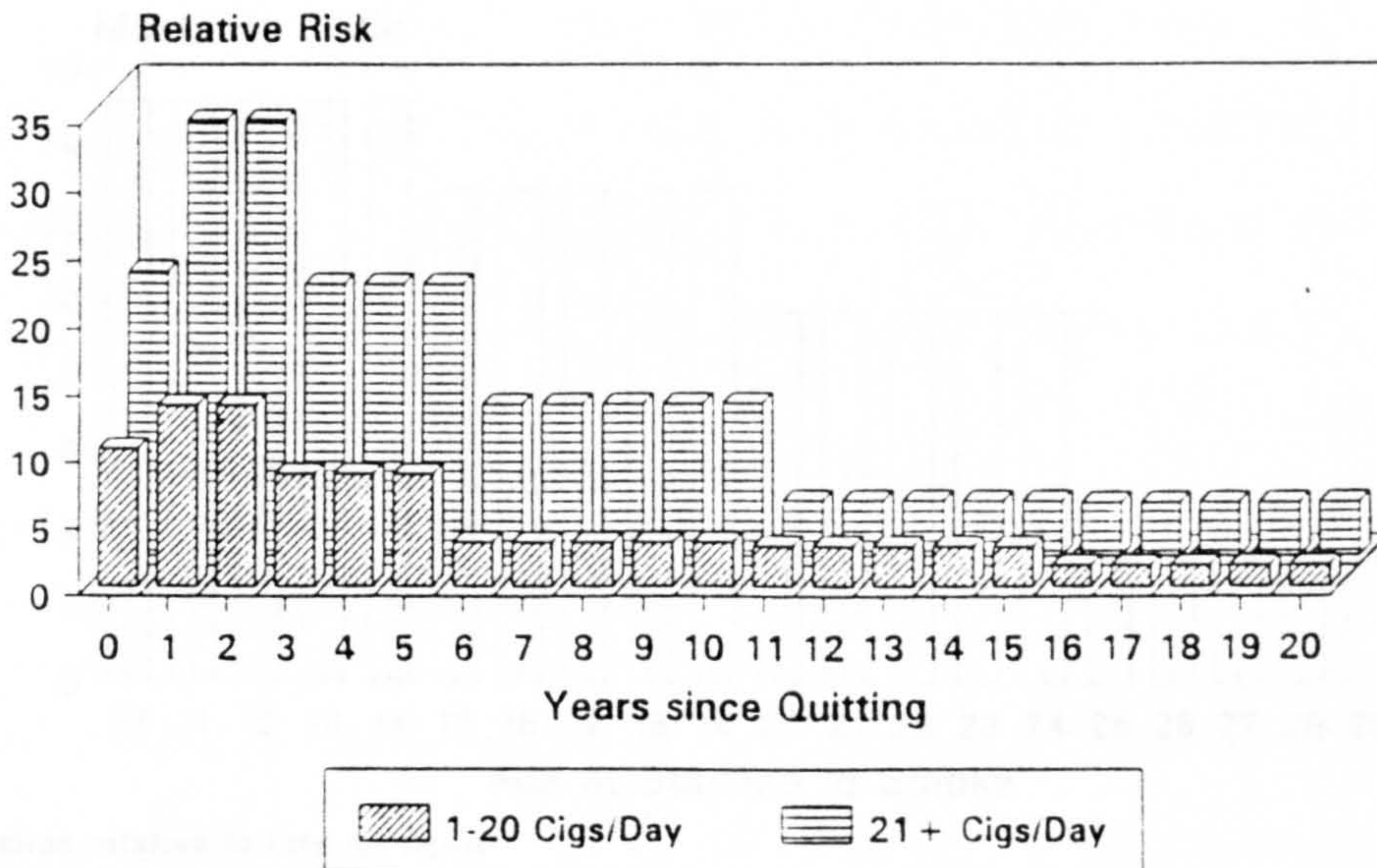


Figure 2.11a Relative Risk for Lung Cancer according to the Time since Quitting, Male



Hammond 1966

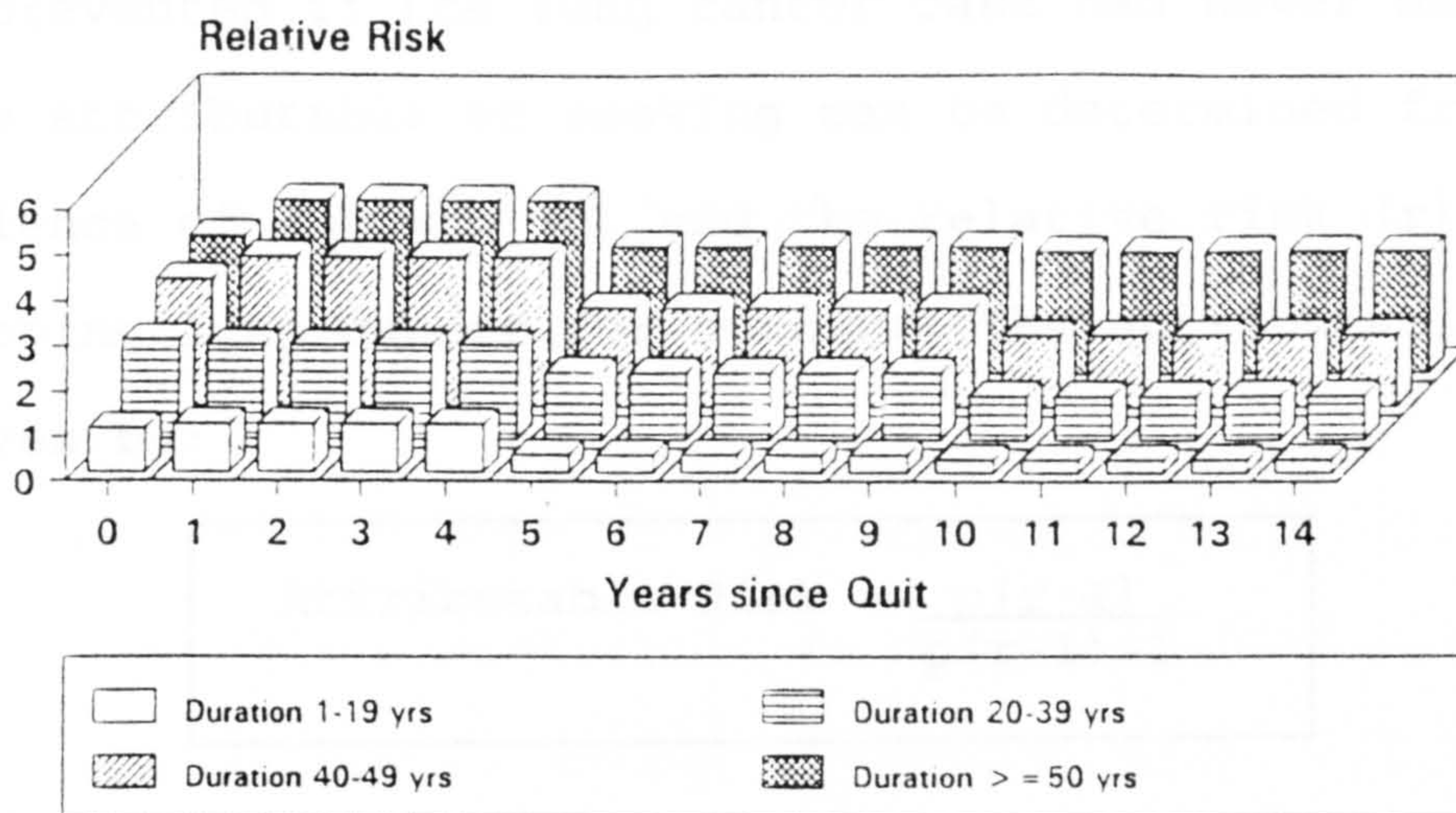
Figure 2.11b Relative Risk for Lung Cancer according to the Time since Quitting, Female



Garfinkel and Stellman 1988

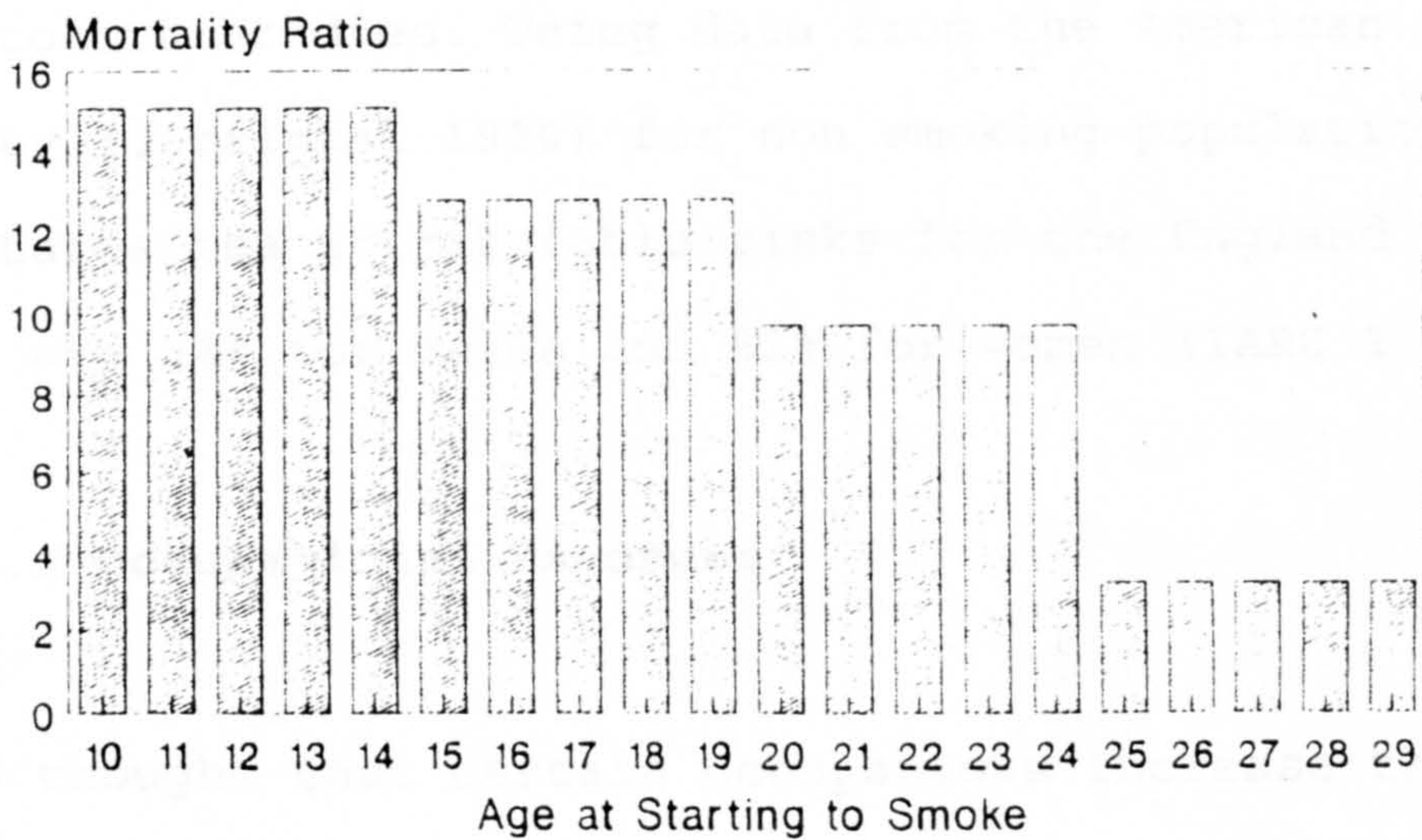


Figure 2.12 Relative Risk of Lung Cancer by Time since Quit and Duration of Smoking, Male



Lubin et al 1984  
Baseline-Current Smokers who had smoked for 0-19 years

Figure 2.13 Mortality Ratios for Lung Cancer according to Age at Starting to Smoke, Male



Ratios relative to rate for never having smoked regularly  
Hammond 1966



The increased risks of lung cancer due to smoking would indicate that at least some of these cancer could have been prevented if the lung cancer case had never smoked. Deaths attributable to smoking can be determined from prevalence of smoking (p) and the relative risk (r) of developing lung cancer for smokers. The attributable risk is given by:

$$\text{Attributable Risk} = \frac{p(r-1)}{p(r-1)+1}$$

Thus for a relative risk of 10 for use as compared to non use of tobacco, and prevalences of smoking of 50% for males and 35% for females, the population attributable risks are 80% for males and 76% for females (Doll and Hill 1976, Doll et al 1980, Hammond 1966). Another approach is to investigate mortality rates in non-smokers from cohort studies. Using data from the American Cancer Society (Garfinkel 1980) for non smoking populations in the States the attributable risks for the England and Wales are 94% for males and 83% for women (IARC 1985).

### 2.3.2 Occupational Exposure

It is thought that certain occupations increase the risk of lung cancer. This is in part due to exposure to certain suspected carcinogens. The occupational factors that have been documented include exposure to asbestos fibres (shipyard workers, construction workers, painters,

welders and miners); crystalline silica (coal miners, stone cutters, painters and glass and ceramic workers); hexavalent chromium (painters, chromate production workers, chromium platers); nickel (welders); beryllium (welders); pitch volatile (aluminum production workers); radon and its decay products (miners) (Tomatis 1990).

In Glasgow and the west of Scotland, where ship building has been a major industry, it is estimated that 5.7% of all male lung cancers during the period 1975-1984 are related to asbestos exposure. The study uses mesothelioma incidence rates as an indicator for asbestos exposure. Lung cancer incidence is dependant on past cigarette smoking (measured by mortality from chronic bronchitis), asbestos exposure, air pollution and a social deprivation index. (De Vos Irvine et al 1993)

This estimate is comparable to that obtained by Pastorino et al (1984) from a case control study, in a highly industrialised area of Northern Italy. They estimate that 7.3% of all cases of male lung cancer could have been avoided if the male population were not exposed to industrial pollutants of asbestos and polycyclic aromatic hydrocarbons.

Miners, in particular uranium miners, are exposed to radon and its decay products, which are implicated in the development of lung cancer (Lubin et al 1995). It has



been estimated that 40% of incident lung cancer among miners in the US is due to radon exposure. The exposure response relationship between cumulative exposure, as measured by Working Level Month (WLM) and lung cancer risk, is significant (Kusiak et al 1993, Morrison et al 1988, Radford et al 1984, Howe et al 1986), and has been shown to be linear in the range of exposures that miners would experience (over 50 WLM) (Lubin et al 1995). The effect of cumulative exposure to radon may be modified by the rate of exposure, age of exposure or duration of exposure (Lubin et al 1995)

Exposure to radon, and its decay products occurs not only in occupational settings, but also at low levels in many homes. In England and Wales the average level of radon present in homes is approximately equivalent to 3 WLM (Chaffey and Bowie 1994). If the linear relationship between exposure and development of lung cancer could be extrapolated to these low levels then exposure to radon in the homes could affect the risk of lung cancer.

However, the role of these low levels of radon exposure is unclear. In recent years several case control studies investigating domestic radon exposure have been reported (Letourneau et al 1994, Alvanga et al 1994, Lees et al 1987, Schoenberg et al 1990). These studies show no consistent pattern of trends between levels of radon exposure and the odds ratios for lung cancer. This may be due to inadequate sample sizes.

In relation to occupation exposures the calculation of population attributable risk is complicated by the multiplicative interaction of many carcinogens with smoking (Whittemore and McMilan 1983, Hornung and Meinhardt 1987, Hammond 1979) The attributable risk for these carcinogens would be greatly reduced by cessation of smoking (Pastorino et al 1984, Pukkala et al 1983).

### 2.3.3 Air Pollution

Air pollution is also thought to be a contributing factor to lung cancer incidence. However, it is difficult to determine its effect in the presence of other contributing factors, and to evaluate the possible interactions between smoking and air pollution. The major studies of air pollution and lung cancer investigate the geographical correlation of the two factors. Urban/rural has been used as an indicator for air pollution , and several studies have shown that lung cancer mortality is higher in urban settings than in rural areas (Mancuso et al 1955, Stocks and Campbell 1950, Levin et al 1960). Muir et al (1987) show a high urban rural ratio in both males and females for North Western England where there has been a major pollution problem. Various studies have attempted to investigate trends in lung cancer mortality and reductions in air pollution (Royal College of Physicians 1970, Lawther and Waller 1978). However the results are difficult to interpret since changes in

smoking patterns and tar content of cigarettes were occurring concurrently. In several studies of air pollution and lung cancer there was an increased risk for males and not for females, suggesting an occupational exposure to the carcinogen (Dean et al 1978, Pike et al 1979). Other studies were conducted among non smokers and found no association with air pollution or its indicator: urban/rural (Doll 1953, Friberg and Cedarlof 1978). Small area studies undertaken in areas where there is a known localised pollutant show increased risk in males living close to the source (Pershagen 1985, Lloyd et al 1985, Smith et al 1987, Kaldor et al 1984). However, Kaldor et al (1984) suggests that the increased risk is due to occupational exposure. From the available epidemiological data it is not possible to quantify the increased risk for lung cancer due to air pollution.

#### 2.3.4 Diet

Consumption of a diet high in vegetables and fruit is thought to be protective against lung cancer and may reduce lung cancer risk by up to 50% (Fontham 1990). Similarly a diet low in fat, particularly saturated fat, is also thought to be protective (Alavanja et al 1993). While smoking is the major risk factor for lung cancer, studies among nonsmokers have shown a protective relationship between dietary beta carotene (Mayne et al 1994) and an increased risk for a diet high in saturated



fats (Alavanja et al 1993). However, nonsmokers tend to have a higher proportion of adenocarcinomas than smokers, who are more likely to have squamous cell carcinomas (Byers et al 1984). The protective effect of beta carotene may occur differentially for different histological types.

The main sources of beta carotene are vegetables and fruit, particularly leafy green vegetables. The inclusion of leafy green vegetables in the diet has been shown to have a protective effect against lung cancer in a cohort of women aged between 55 and 69 years in Iowa (Steinmetz et al 1993). This study included both smokers and nonsmokers and adjusted for smoking variables in the analyses. A similar conclusion is drawn from an ecological study of diet and lung cancer in the South Pacific (Le Marchand et al 1995), where lutein, a carotenoid found in dark green vegetables, was found to be protective. This study also reported that a diet high in cholesterol is associated with increased risk of lung cancer. However, these results may be questionable because of the effect of the very low cancer incidence in Fiji. This low incidence rate may unduly influence the regression even though the consumption of dark green is relatively high. A case control study in China among Yunnan tin miners supported the protective effect of other carotene rich vegetables (yellow and light green vegetables), even though they found no relationship with

dark green leafy vegetables (Forman et al 1992). The authors hypothesise that this lack of relationship is due to the consistently high intake of dark green leafy vegetables in Yunnan Province. In all these studies significant dose response relationships are shown. These studies show that for those people who have a high intake of beta carotene rich foods the risk of developing lung cancer can be reduced to up to one half of those people with a low intake of beta carotene rich foods

### 2.3.5 Social Class

Social class is often indicated as a risk factor for the development of cancer, in both a positive or a negative fashion depending on the site (Rimpela and Pukkala 1987). Although socio-economic status (SES) may be related to lung cancer incidence, it is probably an indicator of some other causative agent. In 1980 the Black Report (Department of Health and Social Security 1980) considered four possible explanations: artifacts; natural or social selection, material or structural reasons and cultural or behavioural explanations. Cancer variations among different SES groups can often be explained by differences in behaviours and exposures. For example smoking is more common in the manual SES groups (Wald and Nicolaidis-Bouman 1991) For industrial exposures it is the blue collar workers who are more likely to be exposed to the potential carcinogens.

### 2.3.6 Summary of Aetiology of Lung Cancer

The literature reviewed indicates that smoking plays the major role in the incidence of lung cancer. Smoking is responsible for 80 - 90% of all lung cancer incidence in males and 60 - 80% in females, worldwide (Peto et al 1992). The other major risk factor is occupational exposure. In industrialised countries it is estimated that up to 10% of lung cancer cases are attributable to occupation exposure, in particular asbestos and radon. The role of air pollution is uncertain but it is thought to be of minor importance (Tomatis 1990). Diet, and in particular the consumption of beta carotene rich foods, may reduce lung cancer risk by up to 50% among those who diet is high in such foods.

There is strong evidence of a dose response relationship between the risk of developing lung cancer and quantity smoked. Other smoking factors that are also related to the risk of developing lung cancer are duration of smoking, age at starting smoking and time since quitting for those who have quit. This relationship between smoking and lung cancer has motivated governments to initiate intervention programs to reduce the percentage of the population who smoke and also to reduce the quantity smoked (Health of the Nation, National Cancer Institute Year 2000 Project, North Karelia Project).



## 2.4 Effects of intervention

This section will discuss the use of projections in the evaluation of intervention programs. The form of intervention that is discussed in this section is related to changing risk behaviours in order to reduce cancer incidence. Screening for disease can also be considered to be an intervention. However, there is no screening program for lung cancer and thus the evaluation of screening programs will not be considered further.

The main thrust of programs to reduce lung cancer incidence has been to target smoking behaviour (Department of Health 1993, Greenberg and Sondik 1986). The main target is to reduce cigarette consumption either through reducing the percent of the population who smoke, and/or to reduce the daily consumption of cigarettes per smoker. The effects of these interventions will depend on the aetiological strength of the intervention, the prevalence of smoking and the role of the tobacco smoke in the multi stage process of carcinogenesis. It has been postulated that smoking affects an early and a late stage in lung carcinogenesis (Hayes and Vineis 1989). Thus part of the effect of stopping smoking should be relatively immediate (due to the late stage carcinogen), while that due to an early stage carcinogen may not be apparent for some time after stopping. For those already exposed to this carcinogen there may be no effect at all. For a late

stage carcinogen the intervention would modify the period effect, while for an early stage carcinogen the intervention would modify the cohort effects.

There have been several studies to investigate the effect of interventions designed primarily to reduce the risk of coronary heart disease. One of the main interventions of these studies is to reduce the amount smoked. The other interventions include improvement of the diet, increase of physical activity, and reduction of hypertension.

(MRFIT (Multiple Risk Factor Intervention Trial Research Group 1982), the Finland North Karelia Study (Puska 1973)). The reduction of the amount smoked should also have an effect on lung cancer incidence among the intervention group. In North Karelia the effectiveness of the intervention program, in relation to lung cancer, is determined by modelling the lung cancer incidence rates in both North Karelia, the intervention area, and Kuopio, the control area (Hakulinen et al 1990). The model includes indicator variables related to age, period, and/or cohort, depending on the cancer site being investigated. Other variables are also included in the model are an indicator variable relating to area, North Karelia or Kuopio, and an integer code relating to the post-program period. This last variable is intended to detect any changes in incidence due to the intervention program. None of the changes in lung cancer incidence could be directly related to the effect of the program.

This is in part due to equivalent decreases in smoking among males, and increases among females in both the intervention area and the control area.

The UK Whitehall study, was a randomised control trial of anti smoking advice (Rose et al 1982). In 1968-70 1445 male civil servants were identified who were both smokers and at high risk of coronary heart disease or chronic bronchitis. These males were randomly assigned to an intervention group who were given active anti smoking advice, or a control group. Both groups were followed for ten years. Although initially the number of cigarettes smoked per day decreased among the intervention group, the quantity gradually increased over the next 9 years. Among the control group the number of cigarettes smoked per day decreased over the ten years of follow up, but was still higher than that in the intervention group. The effect of this behaviour modification on the incidence of lung cancer was small, a difference of 8%, which is not statistically significant. This lack of difference in lung cancer incidence may be due to a relatively short follow up of subjects.

Another intervention to reduce lung cancer incidence that has been evaluated is the supplementation of the diet with beta carotene (The Alpha-Tocopherol, Beta Carotene cancer prevention study group, 1994, Hennekens 1996, Omenn 1996). In both the Alpha-Tocopherol, Beta Carotene



study (ATBC) and the Beta Carotene and Retinol Efficacy Efficacy Trial (CARET) (Omenn 1996) study subjects were those at high risk of lung cancer, either cigarette smokers or those who had had a substantial exposure to asbestos. The subjects for the Physicians' Health Study (Hennekens 1996) were US male physicians who had no previous history of cancer, myocardial infarction, stroke or transient cerebral ischemia. Of the physicians 11% were current smokers and 39% former smokers. None of these studies showed decreased incidence or mortality due to lung cancer for the intervention group. The incidence and mortality due to lung cancer was higher in the intervention group in the ATBC study (incidence relative risk = 1.18, 95% confidence interval 1.03 to 1.36; mortality relative risk = 1.08, 95% confidence interval 1.01 to 1.16), and the CARET study (incidence relative risk = 1.28, 95% confidence interval 1.04 to 1.57; mortality relative risk = 1.17, 95% confidence interval 1.03 to 1.33). In the CARET study the intervention was a joint supplementation of the diet by both beta carotene and vitamin A, and thus it is not possible to determine whether the increase in lung cancer was due to either agent alone or an interaction between the two.

The results of these three studies appear not to support the results of observational studies, where high levels of beta carotene are associated with low lung cancer incidence. They do not support the use of dietary supplementation to reduce lung cancer incidence, and would emphasise the importance of reducing smoking as the

main means of preventing lung cancer.

## 2.5 Summary

The implementation of the Chief Medical Officer's Expert Advisory Group on Cancer (Department of Health 1994) will necessitate knowledge about the number of cancer cases that can be expected to occur in the future. Health Authorities in England are already using such information to plan for the rationalisation of cancer services. Monitoring of the Health of the Nation targets for cancer will also be assisted by utilisation of trend analysis of cancer rates and the related smoking information.

One objective in developing models for projections is to ensure adequate simplicity of the model. The simplest models are based on linear extrapolation over time of some function of the summary rates. For the age-standardised rates the rates are extrapolated over calendar period. Age-specific rates may be extrapolated over both calendar period and birth cohort. There is evidence that lung cancer trends depend on both calendar period and birth cohort effects. These trends can be modelled using age-period-cohort analyses and projected rates calculated from the extrapolations of the period and cohort curves. A major problem with linear extrapolation of past trends in lung cancer incidence is the assumption that the trends in smoking have remained constant. However, with the increasing realisation of the harmful effects of smoking cigarettes, there have been major reductions in smoking in England and other countries.

Models of cancer incidence and mortality have been developed that allow inclusion of trends in risk behaviours. These models allow examination of the effects of changes in trends in risk behaviour. The quality of

the data, particularly of the risk factor data is also of importance. The validity of complex models may be compromised by poor risk factor data. The major risk factor for lung cancer has been shown to be smoking behaviour. The risk of developing lung cancer for an individual is increased with increased quantity smoked, duration of smoking, total quantity smoked and younger age at starting to smoke; and decreased with time since quitting. On a population basis lung cancer incidence is also shown to depend on the percentage of the population smoking.

This thesis will investigate the availability and appropriateness of the smoking information relevant to Mersey Region. Lung cancer incidence is available from the Mersey Cancer Registry. Trends in the age-standardised incidence rates and the age-specific rates will be described. Projections based on linear extrapolations of these trends will be compared with projections based on models incorporating information on risk behaviours. Age-period-cohort models will be developed to describe the trends in lung cancer incidence. These models will be refined by the inclusion of the appropriate risk factor information. However, the models that are developed will include only published smoking data, and not smoking data that has been estimated through interpolation or extrapolation. The inherent problems relating to estimating smoking data from published data will thus be avoided. These models will be compared to models based on extrapolation of trends over calendar time.

The models will be used to determine whether the Health of the Nation targets for Mersey Region will be attained. Service needs will also be estimated from the projections.



Chapter 3

**INFORMATION REQUIREMENTS FOR**

**MODELLING LUNG CANCER**

**INCIDENCE RATES**

This research investigates the relationship between time trends in cancer incidence and trends for the related risk behaviour. From understanding these relationships it is hoped to improve projections of cancer incidence for Mersey Region, the assessment of the potential to achieve the "The Health of the Nation" targets for cancer incidence and the determination of future service needs for lung cancer patients.

In order to undertake this research certain data is necessary. The geographic region has to be accurately defined. Geographic reorganisation of health districts in 1974 has complicated this definition. Information on past and current cancer incidence for this geographic region is necessary. This information is available from the Mersey Regional Cancer Registry, for the years 1951 to 1991. Population figures, by age and sex, are required for each of the years for which there are data on cancer incidence, as well as population projection figures for the years in question. These figures are available for the region from OPCS, and the Mersey Regional Information Unit. The other data that are necessary relate to the risk factors associated with lung cancer. As will be explained in this chapter, the main risk factors that are investigated relate to smoking. Data on the other risk factors, such as industrial exposure, air pollution and diet are scarce.

This chapter will present information on:

---

Definition of Geographic Area

Lung Cancer Data

Population Data

Smoking Behaviour

Service Needs for Lung Cancer Patients

---

For each set of data the following properties will be addressed.

---

Source

Calendar Period covered

Age and Sex Breakdown

Quality

---

### 3.1 Geographic Areas

Prior to 1974, in England and Wales the administrative regions for the Health System were the Hospital Regions. The Liverpool Regional Hospital Board (LRHB) consisted of an aggregation of local authority areas (Table 3.1). Where the boundary of the LRHB divides a local authority area, this local authority area is allocated to the hospital region containing the greater proportion of the population. There were minor boundary changes to the local authority



Table 3.1 Administrative Areas in LRHB

Cheshire (part)	Lancashire (part)	
Birkenhead C.B.	Bootle C.B.	Litherland U.D.
Chester C.B.	Liverpool C.B.	Newton Willows U.D.
Wallasey C.B.	St Helens C.B.	Ormskirk U.D.
	Southport C.B.	Prescot U.D.
Bebbington M.B.	Warrington C.B.	Rainford U.D.
Ellesmere Port M.B.	Crosby M.B.	Skelmersdale U.D.
Hoylake U.D.	Formby U.D.	Widnes M.B.
Lymm U.D.	Golbourne U.D.	
Neston U.D.	Haydock U.D.	Warrington R.D.
	Huyton w Roby U.D.	West Lancs R.D.
Wirral U.D.	Kirkby U.D.*	Whiston R.D.
Chester R.D.		
Runcorn R.D.		
Tarvin R.D.		

\*Kirkby U.D. was formed in 1958

areas in the years 1951 to 1973. The Mersey Regional Cancer Registry covered the LRHB, and parts of North Wales and the Isle of Man (Figure 3.1).

The National Health Service was reorganised in 1974 and the Mersey Regional Health Authority (MRHA) was formed. Although similar to the LRHB, Skelmersdale U.D. and Ormskirk U.D. were transferred to North West Regional Health Authority (NWRHA), while Crewe and Macclesfield were transferred from NWRHA to MRHA (Figure 3.1). The MRHA was divided into 10 District Health Authorities (DHA) (Table 3.2).

Figure 3.1

Map of Cheshire and Lancashire showing LRHB and MRHA

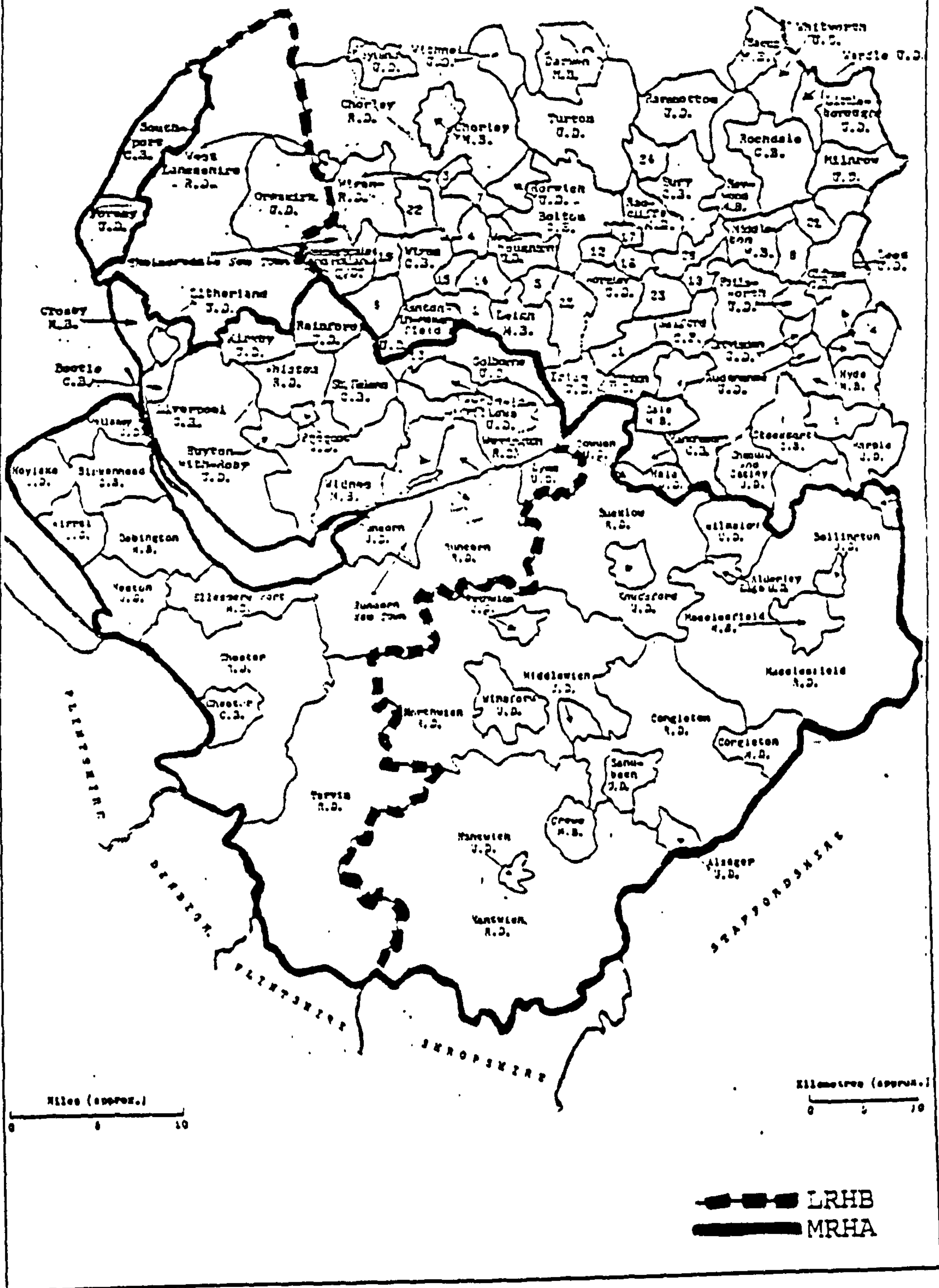


Table 3.2 Health Districts in MRHA

Chester  
Crewe  
Halton  
Macclesfield  
Warrington  
Wirral  
St Helens and Knowsley  
Southport and Formby  
South Sefton  
Liverpool

It is preferable to have a consistent geographic definition of the area under study. Unfortunately this is impossible for the whole time period. Two options that should have been available are:

1. to include the whole area under study ie include Skelmersdale UD, Ormskirk UD, Crewe and Macclesfield for the whole time period 1951 to 1988. This proved impossible because the data for Crewe and Macclesfield are not available from the NW cancer registry prior to 1975, due to re computerisation of the NW cancer registry (personal communication Sandra Gravestock)
2. to include only the common area to both the LRHB and the MRHA. This also proved impossible because the geographic coding prior to 1970 was imprecise with approximately 30% of the cases being coded as "Lancashire in LRHB not otherwise specified" or "Cheshire in LRHB not otherwise specified".



Given the lack of specificity of the geographic codes it is not possible to define the old region in terms of that which is common for the time periods before and after the reorganisation in 1974. Therefore, the decision was taken to analyze data from LRHB for 1951 to 1973, and data from MRHA from 1974 to 1988 (Table 3.3).

Table 3.3 Definition of Geographic Area

1951 - 1973	LRHB
1974 - 1988	MRHA
Future Years	MRHA

### 3.2 Mersey Regional Cancer Registry

The Mersey Regional Cancer Registry attempts to register all incident cancer cases occurring in its jurisdiction. The registry started in 1944 and has computerised records from 1951.

Data were requested from the Mersey Regional Cancer Registry for each registered cases of male lung cancer, and female lung cancer, for the years 1951 to 1987. Lung cancer has a code of 162 in the International Classification of Diseases for Oncology (1976).

Information was requested on the diagnosis, morphology, the data of diagnosis, sex, age and residence of the case (Table 3.4).

Table 3.4 Information Requested on Each Cancer Case

Cryptic patient i.d.  
ICD0 code  
Morphology code  
Month and Year of Diagnosis  
Sex  
Month and Year of birth  
Age at diagnosis  
Postcode  
Area residence code  
Health district

Details of the specific codes for fifth digit morphology codes and residence are given in Appendices III(1) and III(2).

Although the cancer registry holds data from 1951 to 1988 on computer file, it would appear that the completeness of the registry increases dramatically in the early 1950s. This is due to an initiative on the part of the Ministry of Health to increase cancer registration in England and Wales to 100% (Ministry of Health 1954). The LRHB registry was used as a demonstration registry. In order to achieve this goal, three new activities were undertaken. They were:

- the payment of a fee (five shillings per registered case) to all general practitioners registering cancer cases that were not referred for further

treatment. This was initiated in 1956. At that time it was felt by some GPs that a diagnosis of cancer was a death sentence, and therefore referral was not necessary. If a patient was not referred, then s/he could only be registered by the GP. It was felt that this would apply to approximately 1% of cancer registrations

- improvement in the death notification. Previously the registry had been sent only the name and date of death for any death certificate that mentioned cancer. After January 1957 the registry was sent more complete information on all deaths where cancer was mentioned on the death certificate, as well as all those cases registered as having cancer.
- active follow up of hospital records. Previous to the initiative, the individual hospitals were responsible for submitting information on patients diagnosed as having cancer. After January 1958, the registry sent Records Officer's clerks to all the hospital in the region to abstract information on all cancer patients, admitted to hospital.

The effect of these activities can be seen in the increase in registration of lung cancer cases from 713 cases in 1955, 969 cases in 1956 to 1176 cases in 1957, with little change in population structure.



The residence of the patient is identified by one or more of four possible geographic codes (Appendix III(2)).

- the General Register Office Code (4 digits);
- the OPCS code (2 digits and 2 letters);
- postal code (3 characters the first one being a letter followed by up to 4 characters); and
- the Health Authority Code (2 digits).

Between 1951 and 1970 the General Register Office Code was mainly used. This was converted into a dummy postal code which consists of the four digit code inserted into a dummy post code of Q## ##QQ. The Health Authority code was not used at all frequently. From 1970 to 1973 the OPCS code was mainly used. The postal code was used with increasing frequency in this period. Where the postal code was not known a dummy code was created by inserting the OPCS code between Qs as above. Health authority codes were also used in the majority of cases. After 1974 and the reorganisation of the Health Authority, postal codes, health authority codes and OPCS codes were all used. This is the time of the major changes in the boundaries of Mersey Region. It 'lost' Skelmersdale and Ormskirk to the Manchester Region and acquired from the same region Crewe and Macclesfield. (Appendices III(2a and 2b)).

This data set was cleaned, as far as possible to remove those registrations from the Isle of Man and North Wales. The North Wales registrations occurred mainly before the

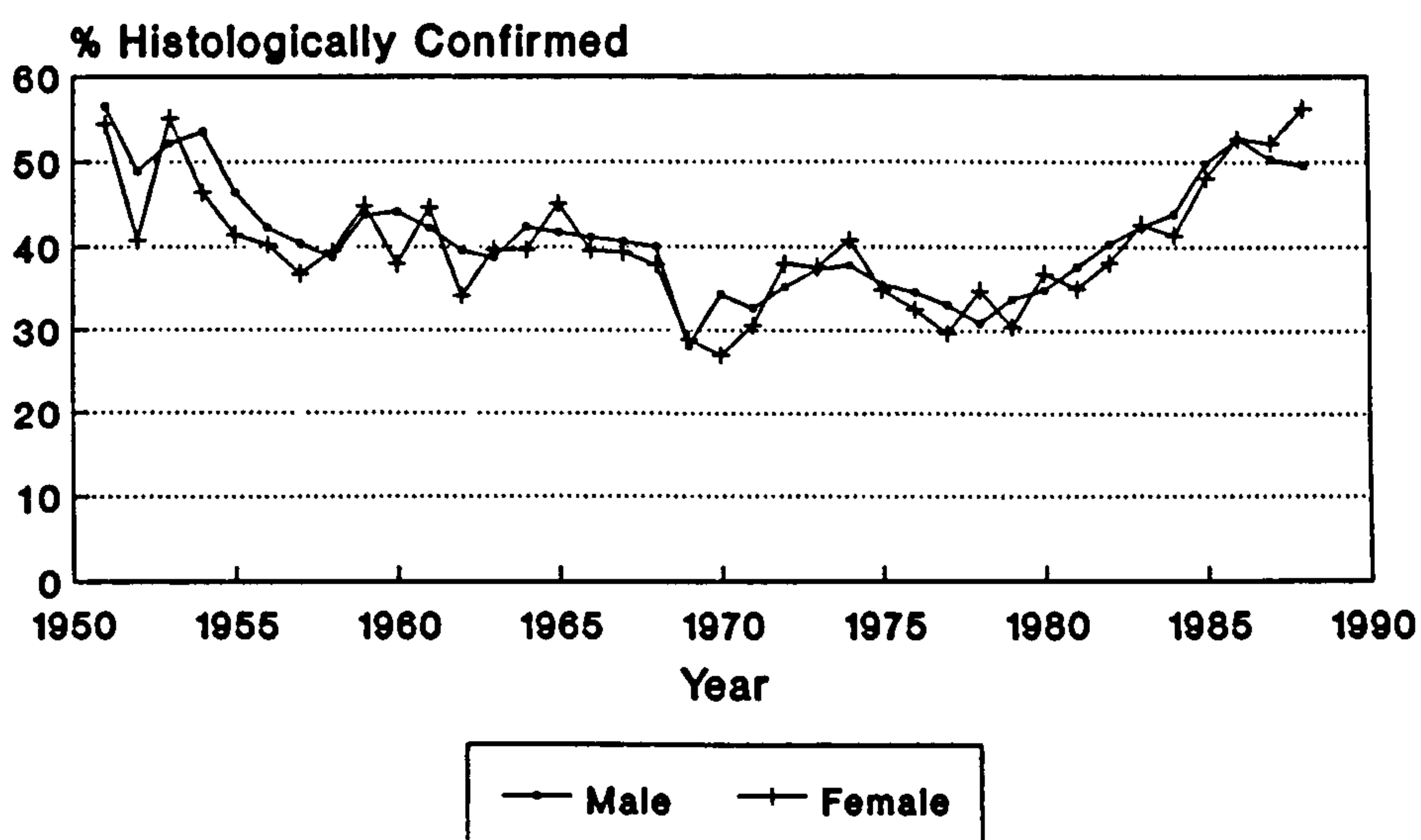
reorganisation in 1974. The Isle of Man data is still collected by the Registry but will not be used in this thesis. Also deleted from the data set are those codes relating to areas known not to be in the Mersey region and for those given an unspecified residence code eg '99\*\*'. Another problem with residence codes is that there was active follow up of patients until 1985. If a patient moved the residence code was updated to the new residence code. I have made the assumption that if a code is outside the region then it will be deleted, since research has shown that the majority of such codes are due to patients coming to Mersey region for a particular treatment that was 'only' available here. (Judith Youngson, personal communication). During this cleaning process, several problems with the data that had occurred at the times of recomputerisation were identified and rectified.

The data were then checked for valid ICD code (162) and Morphology codes. All cases with non malignant lesions were removed; fifth digit morphology codes of 0,1 or 2. The proportion of registered cases of lung cancer with non malignant lesion was small. At present all pathologies are included in the analysis.

If the trends in cancer incidence are to be analysed, it is necessary to ensure the quality of the data is consistent for the whole time period. Two measures of

quality were investigated; the proportion of cases registered through death certificate only and the percentage of registrations that were histologically confirmed. In the recomputerisation of 1987, the information that allowed examination of the percentage of registered cases that were identified through death certificate only (DCO) was lost for the previous years. The percentage of DCOs at present is about 4%. It is thought that this percentage would be representative of the registry, after the late 1950s, because of the use of active follow up of all registered cases (Judith Youngson personal communication). The percentage that are histologically confirmed remained fairly stable, between 30% and 40%, from the late 1950s to the early 1980s (Figure 3.2).

**Figure 3.2 Percent of Registered Lung Cancer Cases Histologically Confirmed, by Sex**



Mersey Regional Cancer Registry



In more recent years the percentage has increased and now over 50% of registered cases are histologically confirmed. The reporting procedures have not markedly changed since the late 1950s, the completeness and quality of the registry appears fairly constant over the last 30 years.

The final data set includes information on all registered cancer cases for malignancies of the male lung, and female lung in LRHB in 1951 to 1973 and in the MRHA in 1974 to 1988.

### **3.3 Population Estimates.**

Population figures for the same geographic areas, and same years as are available for the cancer incidence data are required to calculate the cancer incidence rates. Thus it is necessary to determine the population figures for LRHB for the years 1951 to 1973 and MRHA for 1974 to 1988.

#### **3.3.1 Population Estimates for 1951-1988**

In all cases the 'best' available estimates of the home population have been used. For non census years these are the revised estimates of the home population.

Home population is defined as -

the population of all types, actually in England and Wales distributed by area according to residence.

Population data are available from the Office of Population Censuses and Surveys. Total population censuses are carried out every ten years, with a 10% sample survey carried out in some of the intervening quinquennia. Data are available by local area, sex and five year age group for these years (OPCS 1954a, 1954b, 1964a, 1964b). The Registrar General also produces an annual report that includes intercensal population estimates for defined areas of England and Wales (The Registrar General 1963-1973).

For the census years (1951, 1961, 1971 and 1981) and the years when a 10% sample survey was carried out (1966, 1976 and 1986) data for the local areas are available by 5 year age groupings up to 90-95 years for each sex separately. For the years 1963 to 1973 population data for the LRHB are available for 5 year age groups from 0-4 years to 70-74 years for each sex separately. These data are available from the Registrar General Statistical Review (1963-1973). For the years 1975 to 1988 age specific data are available for each sex separately through the MRHA Information Office. This data originally was provided by OPCS and are based on the 1991 census.

For 1974 the revised home population figures for the MRHA were used.

For the other years there are no intercensal estimates by age and sex for either the local areas or the LRHB.

Therefore a decision was made to linearly interpolate the population figures for each age group, between 1951 and 1961 and between 1961 and 1963. This is of necessity crude but appears to give satisfactory population figures.

### **3.3.2 Populations Projections up to the Year 2011**

Age-specific population projections for each sex separately for Mersey Regional Health Authority for the years 2000, 2001, 2010 and 2011 were obtained from the Mersey Region Information Unit. These figures are based on the OPCS population projections based on 1993 data (OPCS 1995). The projection populations for the MRHA are calculated from projections for the constituent Local Authority Areas. These projections are based on current population and past trends in birth rate, death rate, migration within England and migration outside England. Mersey Regional Health Authority has a declining birth rate and death rate and a declining net emigration.



### 3.4 Risk Factor Information

Population based information on the risk factors for lung cancer are required for the proposed analysis. The major risk factors for lung cancer have been identified as smoking, industrial exposure to carcinogens, possibly air pollution, diet and socioeconomic status. Data on occupation are available for the North West of England (OPCS 1966a, 1966b, 1976, 1984a, 1984b). The data are not adequate for the analyses because the definitions and grouping of the major occupations vary over time, and there is no information on duration of employment. There is little information on trends in dietary habits, particularly the consumption of vegetables and fruits, for the time period of interest. Greaves and Hollingsworth (1966) give some information on consumption of the broad food groups for the United Kingdom, which would not be adequate for the analyses because dietary consumption patterns in Britain vary considerably over geographic regions, and the data are not detailed enough.

It has been estimated that 80 to 90% of all lung cancers are due to cigarette smoking compared to only 10% due to occupational exposures (Section 2.3.5). Although a diet high in beta carotene is estimated to reduce lung cancer risk by 50% compared to a diet low in beta carotene, the overall decrease in lung cancer due to changes in diet is

thought to be small. This is in part due to the average diet in Mersey Region not being high in beta carotene, and in part due to the relative stability of the diet in this region. Since the carcinogenic effect of industrial exposures on lung cancer incidence, is compounded by exposure to cigarettes smoke, and is relatively small, the decision was taken not to include occupational exposure in these analysis. The effects of air pollution and socio-economic status appear only to reflect different smoking habits in the different environments, urban/rural, and in the different socio economic groups (Section 2.3.3). Thus only the role of smoking behaviour in changes in lung cancer incidence will be investigated.

The smoking data that are required for the analysis should be population based and be available from approximately the 1930s. This would enable the analyses to incorporate lag times of up to 25 years between exposure to smoking and the development of lung cancer. For some of the risk factors it is important to know the cohort exposure, for example quantity smoked.

Given the requirements above the smoking parameters that were investigated were:

1. the number of manufactured cigarettes smoked per person per year
2. the percent of the population who smoke cigarettes
3. the number of manufactured cigarettes smoked per smoker per week
4. the percent of smokers who smoke plain cigarettes
5. the average tar yield per cigarette

There is no suitable information on the percentage of people quitting and thus this was not included in the analyses.

The most important source of smoking information is "U.K. Smoking Statistics, second Edition" (Wald and Nicolaides-Bouman 1991) The information in this book consists of a amalgamation of published and unpublished data from the General Household Survey (GHS) and the Tobacco Advisory Council (TAC). These data mainly relate to either Great Britain (England, Wales and Scotland) or to the United Kingdom (Great Britain and Northern Ireland). There is little historic data on smoking that relates to Mersey Region exclusively. Therefore national data on smoking are used for the analyses. The source of the data on smoking used in these analyses is the TAC. The TAC data on quantity smoked and the percentage of the



population who smoke are determined from annual surveys on smoking habits in Great Britain. Approximately 10,000 people are interviewed for each survey. Quota sampling is used. The sample is stratified by sex, age, social class, region and occupation. The data are adjusted against sales figures to correct for under reporting of cigarette consumption by respondents.

The TAC define a **smoker** as a person who answers **yes** to at least one of the following questions:

Do you smoke packeted cigarettes?

Do you smoke hand rolled cigarettes?

Do you smoke a pipe?

Do you smoke as much as one cigar or miniature cigar a week?

Cigarette consumption was estimated from answers to the following question:

How many cigarettes did you smoke yesterday? (On Mondays, twice the usual number of people are interviewed and half of them are asked 'How many cigarettes did you smoke on Saturday?')

In all the analyses that follow the data refer only to smokers of manufactured, or packeted cigarettes, ie only those people who answered 'yes' to the first question above.

The average number of manufactured cigarettes smoked per person per year in the UK is available from 1905 to 1987 for males and 1920 to 1987 for women (Wald and Nicolaidis-Bouman 1991, Table 2.1). The age-specific cigarette consumption per person per week for Great Britain for males and females separately is also available (Wald and Nicolaidis 1991, Tables 4.9.1 and 4.9.2). The age groups that are presented vary over time (Table 3.5). It is necessary to estimate the consumption for consistent age groups if trends over time are to be examined.

The percentage of the population of Great Britain smoking manufactured cigarettes is available, on an annual basis for the years 1948 to 1987 (Wald and Nicolaidis-Bouman 1991, Table 3.3) Data on the age-specific percentages of the population of Great Britain who smoke are available for each sex separately for the years 1948-1987 (Wald and Nicolaidis-Bouman 1991, Tables 4.1.1 and 4.1.2). The percentages for the age group 34-59 for the years post 1975 have to be approximated (Section 5.1).

Data on the average number of cigarettes smoked per smoker per week in Great Britain is available from 1948 to 1987 for males and females separately (Wald and Nicolaidis-Bouman 1991, Table 2.3). Data on age specific consumption of cigarettes smoked per smoker per week are available for each sex separately for the years 1948-1987

(Wald and Nicolaides-Bouman 1991, Tables 4.10.1 and 4.10.2). As with the number of cigarettes smoked per person, it is necessary to estimate the age-specific consumption for consistent age groups over time (see Section 5.1)

Data on the percentage of smokers smoking plain cigarettes in Great Britain are available for each sex from UK Smoking Statistics (Wald and Nicolaides-Bouman 1991, Tables 3.1.1 and 3.1.2). Those smokers who did not have a usual brand were not included in these figures. However the assumption was made that the percent of those smokers who smoked plain cigarettes would be the same as those who did have a usual brand. These data are available for 1958, 1961, 1963, 1965, 1968 and all the years between 1971 and 1987.



Table 3.5 Summary of Smoking Data Available

Variable	Range	Sex	Age Groups
No. of Cigarettes Smoked per Person per Year	1905-1987	M and F	
	1948-1955	M and F	25-29, 30-34, 35-59
	1956-1975	M and F	25-29, 30-34, 35-49, 50-59
	1975-1987	M and F	25-34, 35-49, 50-64
Percentage of Population who Smoke	1948-1975	M and F	25-34, 35-59
	1975-1987	M and F	25-34, 35-49, 50-64
No. of Cigarettes Smoked per Smoker per Week	1949-1955	M and F	25-34, 35-59
	1956-1975	M and F	25-34, 35-49, 50-59
	1975-1987	M and F	25-34, 35-49, 50-64
Percentage of Smoker who Smoke Plain Cigarettes	1958	M and F	
	1963		
	1965		
	1968		
Annual Sales of Filter and Plain Manufactured Cigarettes	1928-1987	both sexes combined	
Average Tar Content of Cigarettes (sales adjusted)	1934-1969	(7yr average)	
	1970-1988	both sexes combined	

Data are available on the sales adjusted average tar content of cigarettes (mg/cig) sold in the UK (Wald and Nicolaides-Bouman 1991, Tables 8.4). These averages are calculated by weighting the tar content of each brand of cigarettes by the percentage of market share of that brand. The data are based on analysis of the tar content of cigarettes by the Laboratory of the Government Chemist for the years 1934-40 to 1979 (Wald et al, 1981) and from 1980 calculated from The Laboratory of the Government Chemist surveys (Health Departments of the United Kingdom 1973-82), with the market share data from an annual anonymous tabulations entitled "Brand Shares of the UK cigarette market" in Tobacco. For the years 1934 to 1969 the data were available only as averages for seven year periods eg 1934-40. No adjustment was made for differences between sexes because such data are only available from 1978.

### **3.5 Service Needs Information**

The therapy patterns for lung cancer patients for the years 1983-87 are documented in the Lung Cancer Bulletin (Williams et al 1993). The number of cases receiving surgery, chemotherapy and radiotherapy on a yearly basis can be determined for each sex and age group separately (Appendix III(3)). In using these figures to determine service needs in the future, the assumption is made that the age-specific pattern of therapy will not change.

### 3.6 Summary

This thesis investigates the relationship between the trends for lung cancer incidence in males and females in Mersey Region. In order to undertake the analyses, it is necessary to define the geographic region under study, extract information on all lung cancer cases incident in the given region, determine the appropriate risk factors and extract the relevant risk factor data.

The period of study under investigation is 1951 to 1988, being the period for which the Mersey Regional Cancer Registry has computerised data. During this period, the Health Authorities have been reorganised, from Regional Hospital Boards to Regional Health Authorities. Given the difficulties in defining a common area for the whole time period, it was decided to define the area under study as the LRHB from 1951 to 1974 and the MRHA after 1974.

Lung cancer incidence data are available from the Mersey Regional Cancer Registry for the years 1951 to 1988. Data on diagnosis, date of diagnosis, age, sex and residence are available for each case of lung cancer registered within the defined areas (Table 3.6). Population data are available from the OPCS. Service needs data are available from the Lung Cancer Bulletin (Williams et al 1993).



**Table 3.6 Information Sources for Data relating to Lung Cancer Incidence, Population and Service Needs**

**Geographic Area**

1951-1973 LRHB

1974-2011 MRHA

**Lung Cancer Incidence**

Source: Mersey Regional Cancer Registry

Period: 1951-1988

Data: Information on each cases registered in the geographic areas, including:  
date of birth,  
sex,  
year of diagnosis,  
morphology and  
area of residence

**Population Data**

Source: OPCS, Registrar General's Statistical Review, Mersey Region Information Office

Period: 1951-2011

Data: Home populations for Mersey Region for five year age groups for males and females separately

**Information on Service Needs**

Source Mersey Regional Cancer Registry - Lung Cancer Bulletin 1993

Period 1983-87

Data Numbers and percentages of procedures by age and sex

The major risk factor for lung cancer is smoking behaviour. This is thought to account for 90% of all lung cancer incidence. National data are available for measures of quantity smoked, the percentage of the

**Table 3.7 Information Source for Data on Risk Factors for Lung Cancer**

**Information on Smoking Behaviour**

Source: Tobacco Advisory Council (Wald and Nicolaides-Bouman 1991)

Data: Number of Cigarettes Smoked per Person by Sex: 1905-1987

by Age and Sex: 1948-1987

Percentage of the Population who Smoke by Age and Sex: 1948-1987

Number of Cigarettes Smoked per Smoker by Age and Sex: 1949-1987

Percentage of Smoker who Smoke Plain Cigarettes

By Sex: 1928-1987 (estimated from the total annual sales of filter and plain cigarettes)

Average Tar Content of Cigarettes: 1934-1988

**Information on Occupational Exposure**

Source	Data of sufficient quality not available
--------	--

population smoking, and the composition of cigarettes for sufficient years to allow different lag times to be investigated. Adequate smoking data for Mersey Region are not available, and therefore it was decided to use the national figures (Table 3.7). Other risk factors, such as occupational exposure, were not investigated, because of paucity of the data, and the relatively small impact they have on lung cancer incidence.

Chapter 4

**TRENDS IN LUNG CANCER INCIDENCE**

**IN MERSEY REGION AND**

**SIMPLE METHODS OF PROJECTION**



This chapter investigates the trends in lung cancer incidence in Mersey Region for the period 1951 to 1988. Projected lung cancer incidence rates will also be described. The trend analysis will:

- 
- describe the trends in the total annual number of incident lung cancer cases,
  - describe the trends in the crude lung cancer incidence rates,
  - determine the standard population to use in the examination of the age-standardised lung cancer incidence rates
  - describe the trends in the age-standardised lung cancer incidence rates
  - describe the trends in the broad-band age-specific incidence rates, age groups 30-44 years, 45-64 years and 65+ years
  - describe the trends in the 5 year age-specific incidence rates over calendar period
  - describe the trends in the 5 year age-specific incidence rates over birth cohorts
  - determine if there are any differences between trends for males and females
  - determine if the trends are due to:
    - changes in registration practise
    - changing population structure
    - calendar period or birth cohort effects
- 

The analyses of the trends in lung cancer incidence will inform the decisions about the preferred methods for projecting lung cancer incidence for the years 1990, 2000, and 2010.

The section on projections will:

---

- determine what standard population to use for projections based on the age-standardised lung cancer incidence rates
  - calculate projected lung cancer incidence rates based on linear extrapolation of the log of the:
    - the age-standardised incidence rates
    - the broad-band age-specific incidence rates
  - determine on which range of years should the projections be based
  - determine how the projections for 1990 compare to the number of cases registered for that year
  - compare the projections based on the two methods
  - determine if the Health of the Nation targets will be met
  - determine service needs for the year 2000
- 

In this thesis, most graphs of incidence rates are presented in the natural scale. In some instances, the logarithmic scale would be more appropriate for the interpretation, and in these cases the graphs are presented using logarithmic scaling.

## 4.1 Trends in Lung Cancer Incidence Frequencies

### 4.1.1 Methods

The total number of registered incident lung cancer cases for each sex and year, are determined from the Mersey Regional Cancer Registry data. These data are plotted against year to determine the trends over time.

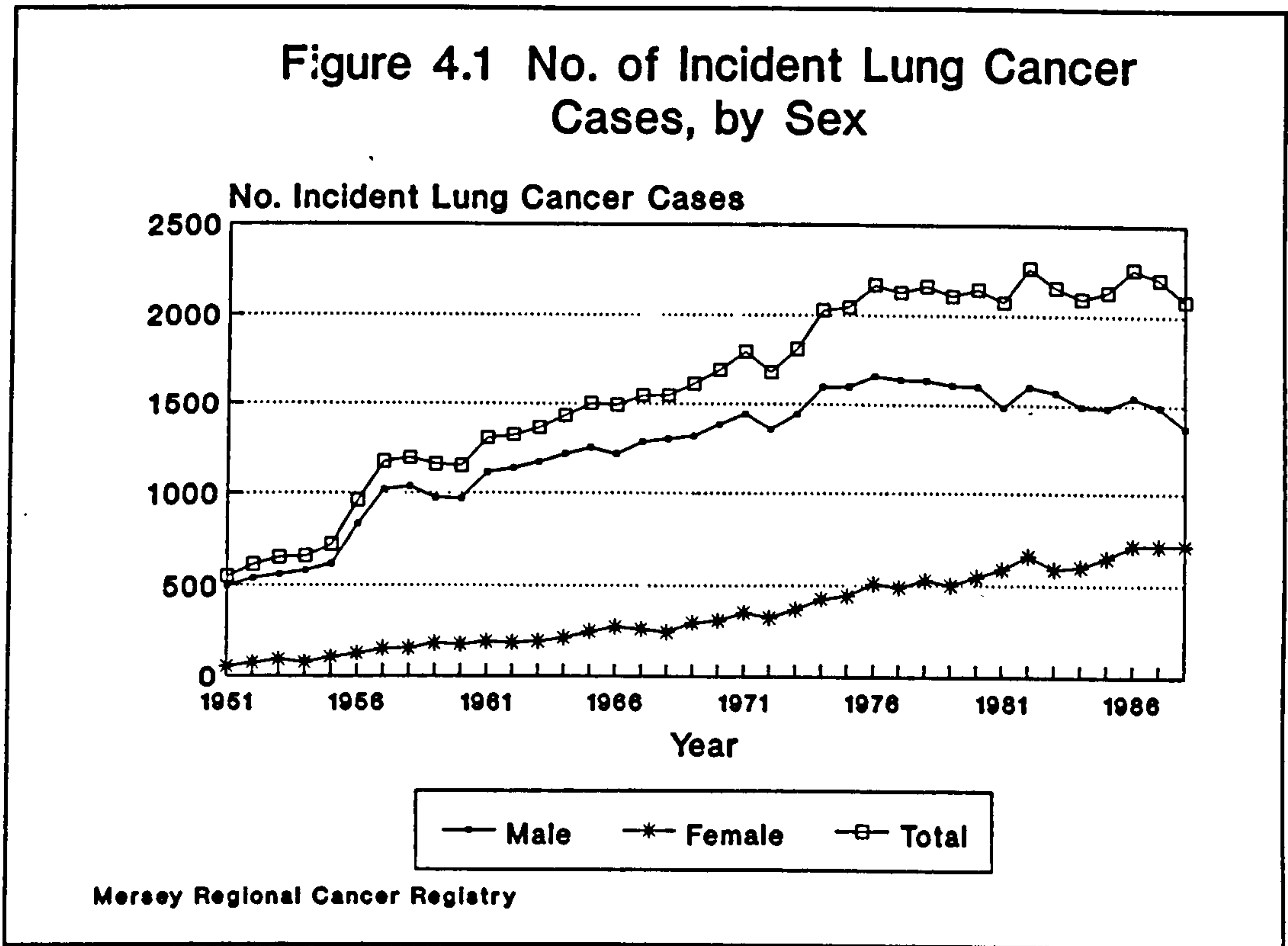
### 4.1.2 Results

The annual number of incident lung cancer cases registered for males increases from approximately 500 cases in 1951 to over 1600 cases in the mid 1970s (Figure 4.1). The increase is most dramatic in the mid 1950s. This rapid increase was due in part to the initiative on the part of the central government to increase cancer registration to 100% in the LRHB during this period. After 1980 the number of cases registered each year begins to decline to around 1400 in the late 1980s. This decline is thought to be a decline in incidence and not a change due to registration practices.

In 1951 only 52 cases of lung cancer were registered for females (Figure 4.1). Since then the number of lung cancer cases registered per year has increased steadily to around 700 in the late 1980s. Thus, in the late 1980s the annual number of incident lung cancer cases in



females was approximately half that in males.



The total number of registered lung cancer cases in Mersey region increased from around 500 in 1951 to over 2000 by 1976 (Figure 4.1). After that time the total number has remained stable, with the decrease in males being compensated for by the increase in females.

## 4.2 Trends in Crude Lung Cancer Incidence Rates

### 4.2.1 Methods

Although the number of incident cancer cases is of importance in determining service requirements, this number may depend on the size of the population at risk.

If this population is increasing or decreasing over time any increase or decrease in the total number of cancer cases may be due only to this change in population size, and not to changes in risk behaviours or other factors.

The geographic region that is used for this study has been defined as the LRHB prior to 1974, and the MRHA for 1975 onwards (Section 3.1). From 1951 to 1974 the population of the LRHB was increasing for both males and females; from 1.00 million to 1.07 million males and from 1.09 million to 1.14 million females. In 1975 the population for the MRHA was 1.21 million males and 1.30 million females. These populations have decreased to 1.17 million males and 1.24 million females in 1988. The changes in population sizes would affect the number of incident cancer cases in the region even if the underlying incidence rates remained constant (Registrar General 1965-1975, MRHA Information Office 1975-1988, OPCS 1954, 1964).

The simplest means of adjusting for changes of population size is to calculate the crude rate for a given time period, where:

Crude Rate/Time Period

$$= \frac{\text{No. of Incident Cancer Cases in a Time Period}}{\text{Person Years at Risk for the Same Time Period}}$$

The time period that is used for the analyses of the trends in the crude lung cancer incidence rates is one year. Person years at risk is estimated by the mid-year population for a given year times 1 year.

Crude lung cancer incidence rates are calculated for each year from 1951 to 1988 for both male and female lung cancer incidence. These rates are plotted against year. The trends are discussed in relation to changes in the population size.

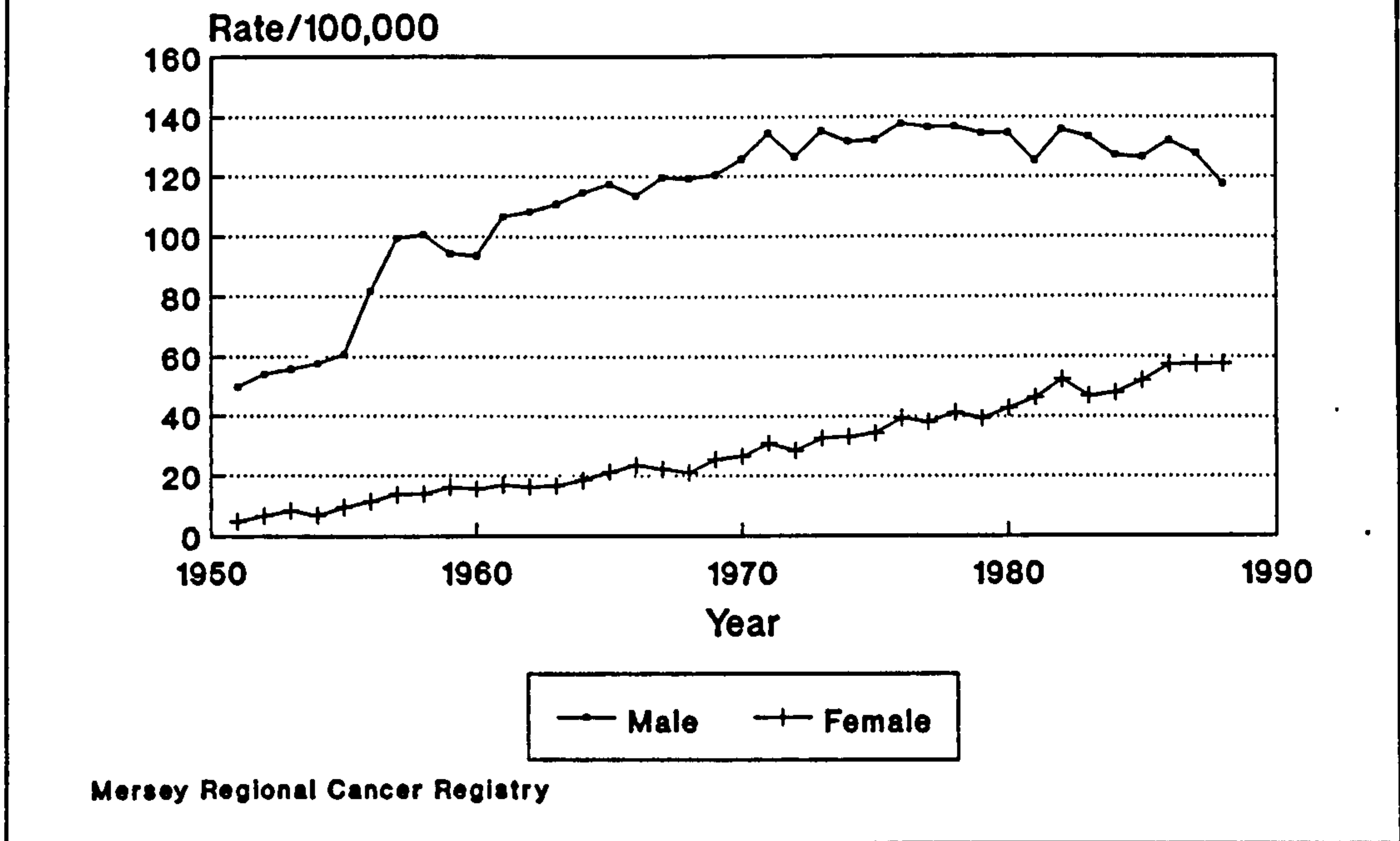
#### 4.2.2 Results

The crude lung cancer incidence rates increased for both males and females over the study period (Figure 4.2). For males the rates increased from approximately 50 per 100,000 in 1951 to just over 100 per 100,000 in the early 1960s, with a rapid increase from 1955 to 1957, during the initiative to improve cancer registration. During the 1960s the rates increased more slowly to around 135 per 100,000. During the 1970s it maintained this level and then began to decrease slowly in the 1980s to around 120 per 100,000.

The crude lung cancer incidence rate for females increased steadily over the entire period from around 5 per 100,000 in the early 1950s to just under 60 per 100,000 in the late 1980s (Figure 4.2).



Figure 4.2 Crude Lung Cancer Incidence Rates, by Sex and Year



In the early 1950s the female crude lung cancer incidence rate was approximately 10% of the male rate. This increased to 17% in 1960, and 19% in 1970. The increase then became more rapid with the female rate being 30% of the male rate in 1980, and in 1988 the female crude lung cancer incidence rate was approximately half that for the males (Figure 4.2)

Thus the patterns for the crude lung cancer incidence rates are similar to that for the frequencies of lung cancer for both males and females. However, the changes in population size for both males and females lead to the decrease in male crude lung cancer incidence rates being less marked than that for the frequency, while for females the increase in crude lung cancer incidence rates

being more marked than that for the frequency.

#### 4.3 Calculation of Age-Specific Lung Cancer Incidence Rates

Although the crude incidence rate accounts for changes in population size it does not adjust the rate for changes in the population age structure. Lung cancer is a disease of the elderly (Muir et al 1987), and changes in the proportion of the population in the older age groups, (eg over the age of sixty five), would have an effect on the crude incidence rates. Increases in the proportion of the population over the age of sixty five would increase the absolute number of lung cancer cases, without there necessarily being any change in the underlying incidence rate. The male population age structure has been changing over the period of the study. Between 1951 and 1988 the percentage of the population that is over 65 years of age has increased steadily from 8% to 12% for males. This would indicate that the decrease in the crude rates for males in the 1980s is due to falling incidence rates. The increase in the female crude rate may be explained in part by the increase over the same period, of the proportion of the female population aged over 65 years, from 11% to 18%.

In order to account for the changing age structure of the population, age-specific rates can be used. The age-

specific rate is the crude rate for a specified age group. The most frequently used age intervals for age-specific rates are five year intervals; e.g. 30-34 years, 35-39 years, etc. These five year age-specific rates will be used in this thesis for plotting, calculation of age-standardised rates and age-period-cohort modelling. Other age intervals will also be used in this thesis; i.e. 30-44 years, 45-64 years and 65 years and above. These latter ones will be used for the calculation of broad-band age-specific rates.

Age-specific rates are calculated as:

$$\text{Age-specific Rate for age group } (a, a+A-1) / \text{time period} \\ = \frac{\text{No. of cases in age group } (a, a+A-1) \text{ in a time period}}{\left[ \begin{array}{l} \text{person years at risk in age group } (a, a+A-1) \text{ for} \\ \text{the same time period} \end{array} \right]}$$

where  $a$  is the lower age limit

and  $A$  is the width of the interval

$$\text{or } R_i = I_i / N_i$$

where  $i$  indexes the age group

$R_i$  is the age-specific rate for age group  $i$

$I_i$  is the number of incident cancer cases in age group  $i$ , and

$N_i$  is the person years at risk in age group  $i$

The time periods used in this thesis are either one year or five years (quinquennia). Age-specific rates using a one year time period give a detailed description of the



situation, but are often too complex to enable a simple interpretation of the whole picture. Also, because they may be based on small numbers of cases they may subject to considerable variation. Use of summary measures of age-specific rates, such as age-standardised rates, overcomes these problems but leads to loss of information regarding changing relationships of incidence rates with age over time. If five year time periods are used the problems of small numbers also may be overcome, but information will be lost on changes in rates within the five year periods.

**Table 4.1 Age Specific Rates per 100,000 for Male Lung Cancer for LRHB (1959-73) and MRHA (1974-88)**

Age Group	Quinquennia					
	1961	1966	1971	1976	1981	1986
30-34	7.0	4.4	4.9	3.6	2.5	1.5
35-39	11.9	13.7	11.4	9.0	9.3	7.2
40-44	37.7	35.2	36.0	23.2	19.7	15.9
45-49	79.4	85.2	88.7	76.9	60.0	44.8
50-54	161.4	155.5	159.4	159.1	135.8	99.6
55-59	306.7	278.1	289.2	286.7	247.2	213.4
60-64	465.8	484.1	493.8	451.4	420.6	385.1
65-69	562.0	663.8	692.7	644.0	551.2	593.1
70-74	604.1	766.8	833.3	790.4	818.9	708.9
75-79	513.3	744.5	784.9	968.2	929.3	846.7
80-85	360.0	668.6	665.9	901.1	1035.8	956.7

Mersey Regional Cancer Registry

In this thesis five year age-specific frequencies are

Table 4.2 Age Specific Rates per 100,000 for Female Lung Cancer for LRHB (1959-73) and MRHA (1974-88)

Quinquennia	1961	1966	1971	1976	1981	1986
Age Group						
35-39	4.8	5.5	6.3	3.3	5.1	3.0
40-44	10.6	12.5	12.2	11.5	10.7	9.5
45-49	17.1	19.5	23.9	29.8	21.6	26.2
50-54	27.1	38.1	49.9	54.9	50.5	46.5
55-59	35.3	48.2	62.2	85.5	100.9	101.7
60-64	47.4	50.0	79.0	118.0	141.9	180.0
65-69	53.0	67.3	101.6	122.1	157.8	203.3
70-74	57.2	79.8	102.8	127.0	172.4	211.2
75-79	69.2	77.3	103.7	131.4	150.7	194.1
80-85	71.4	97.4	99.0	98.4	153.0	178.1

Mersey Regional Cancer Registry

calculated for the quinquennia centred on the years 1961, 1966, 1971, 1976, 1981, and 1986. In 1961, 1971 and 1981 100% census data are available for the geographic area. In 1966, 1976, and 1986 10% samples of the population are taken. Person years at risk are estimated by multiplying the five year age-specific population frequencies for the given years by five. Five year age-specific incidence rates for each quinquennium are then calculated for both sexes (Tables 4.1, 4.2)

This chapter will use four methods to examine age-specific rates; viz age-standardised rates and broad-band

age-specific rates based on one year time periods; and age-period and age-cohort plots based on the given quinquennia. Each of the methods is now described highlighting their strengths, weaknesses and appropriateness.

#### 4.4 Trends in Age-Standardised Lung Cancer Incidence Rates

Age-standardised rates are used to compare rates over time when the age structure of the population is changing. They are also used when rates are compared among different geographic regions.

##### 4.4.1 Methods

Two methods of standardisation are available; direct or indirect (Fleiss 1981). The direct method of standardisation requires the age-specific rates of the population of interest to be known, and these are applied to the standard population, to determine the overall age-standardised rate.

$$\text{Age Standardised Rate} = R^s_i = \frac{\sum_i (N^s_i \times R_i)}{\sum_i N^s_i}$$

where  $N^s_i$  is the person years at risk for age group  $i$  for the standard population

and  $R_i$  is the age specific rate for age group  $i$  for the population of interest



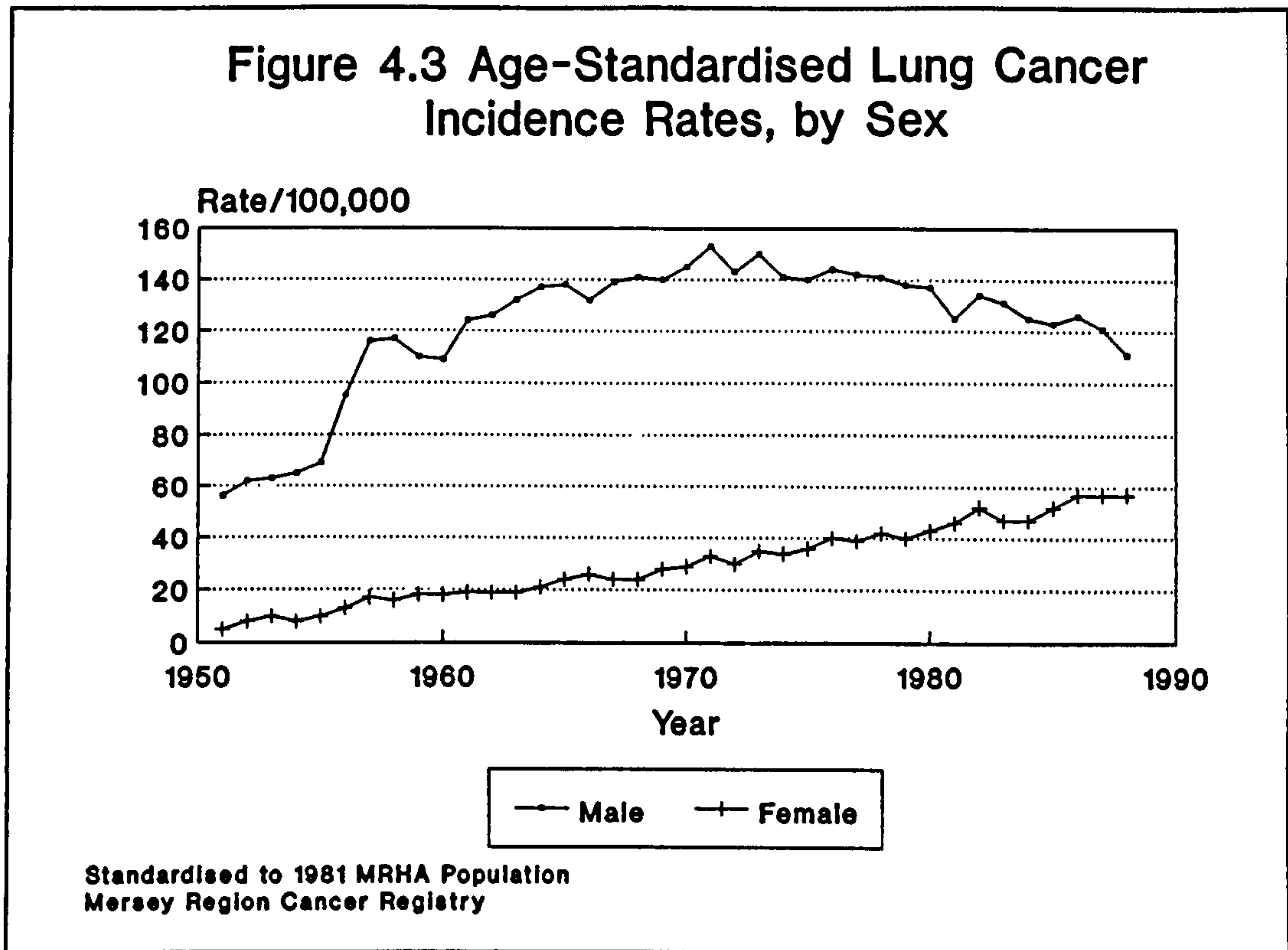
For the indirect method, standard rates are applied to the local population. The main aim of this thesis is to investigate trends in lung cancer incidence in Mersey Region and therefore only direct standardisation will be used. The directly age-standardised rate is the crude rate that would have been obtained if the age-specific incidence rates are applied to the standard population.

The choice of the standard population is important. It should reflect a representative population for which the standardised rates are being calculated. In order to compare the cancer incidence rates over the time period of 1951 to 1988, age-standardised rates, for Mersey Region are calculated for each year. Since the comparisons are across time within the region it is sensible to use as the standard population a region population for a year within the time span. The 1981 population for Mersey Region was used as the standard because there was a full census in 1981, and data are available for Mersey Region. These age-standardised annual rates are then plotted against year.

#### **4.4.2 Results**

The age-standardised rates for male lung cancer increased from below 60 per 100,000 in 1951 to a high of around 140 per 100,000 in 1964. The curve then plateaued until 1980, and then the age-standardised incidence rates declined to

110 per 100,000 in 1988 (Figure 4.3)



As with the crude incidence rates there was a rapid increase in the age-standardised rates between the years 1955 and 1957, from 69.5 per 100,000 to 115.6 per 100,000.

However, the age-standardised rates for males plateau at a higher level than the crude rates (approximately 140 per 100,000 compared to approximately 130 per 100,000 for the crude rates). Also the decline in the more recent years is more rapid. These differences reflect the changing age structure of the population.

The age-standardised rates for female lung cancer are much lower than those for male lung cancer, ranging from

5.4 per 100,000 in 1951 to 58.2 per 100,000 in 1988

(Figure 4.3). As with the males there is an increase in rates in the mid 1950s. Again this is most likely due to the increased registration during that period.

However, in contrast to the males there has been a steady increase in rates for females between 1951 and 1988, with no plateauing in any period. On the log scale this increase appears linear indicating an exponential rise in female lung cancer incidence over the last thirty years (Appendix IV(1)). While the age-standardised rate for females was only 14% of that for males in the late 1950s, by the late 1980s it is almost 50% of that for males.

#### **4.5 Trends in Broad-Band Age-Specific Lung Cancer Incidence Rates**

The next method that is presented is the examination of trends in the broad-band age-specific incidence rates, for each year, 1951 to 1988, for both male and female lung cancer. These age groups 30-44 years, 45-64 years, and 65+ years are chosen to reflect the patterns in lung cancer incidence amongst the young, the middle aged, and the elderly.

##### **4.5.1 Methods**

Broad-band age-specific rates are calculated for the age



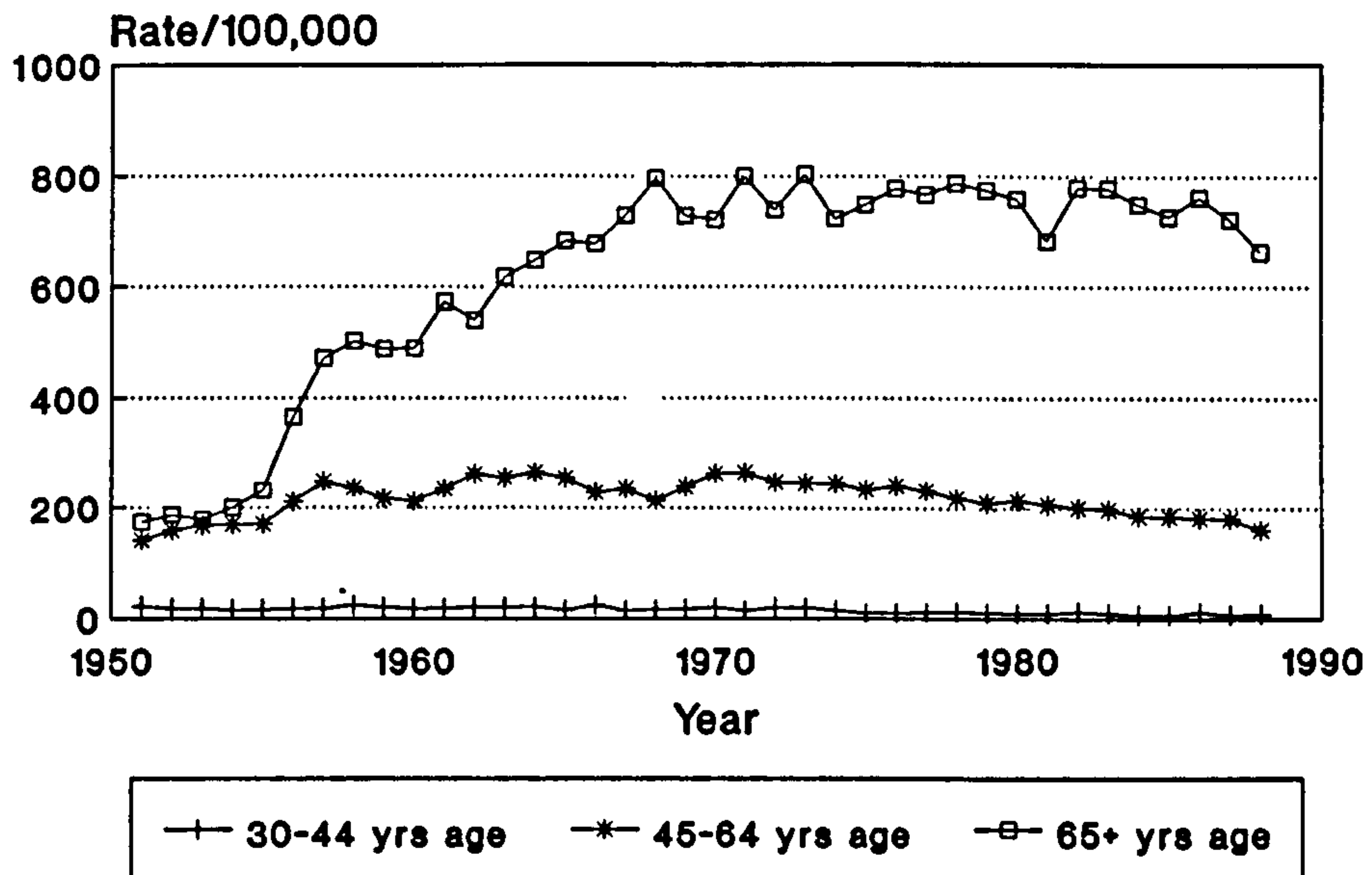
groups 30-44 years, 45-64 years and 65+ years, for each sex separately. The rates are not standardised within these broad age groups, because the variations in the rates within these age bands are not large. These broad-band age-specific rates are plotted against year.

#### 4.5.2 Results

The broad-band age-specific incidence rates for male lung cancer for the 30-44 year age group (the young group) was relatively stable for the first 22 years of the registry, varying between the limits of 15 and 20 per 100,000 (Figure 4.4) In the early 1970s the rate fell substantially to around 10 per 100,000 where it has remained for most of the 1980s. For the 30-44 year old females the rate varied between 1 per 100,00 and 10 per 100,000, showing no consistent increase or decrease for the whole period (Figure 4.5).

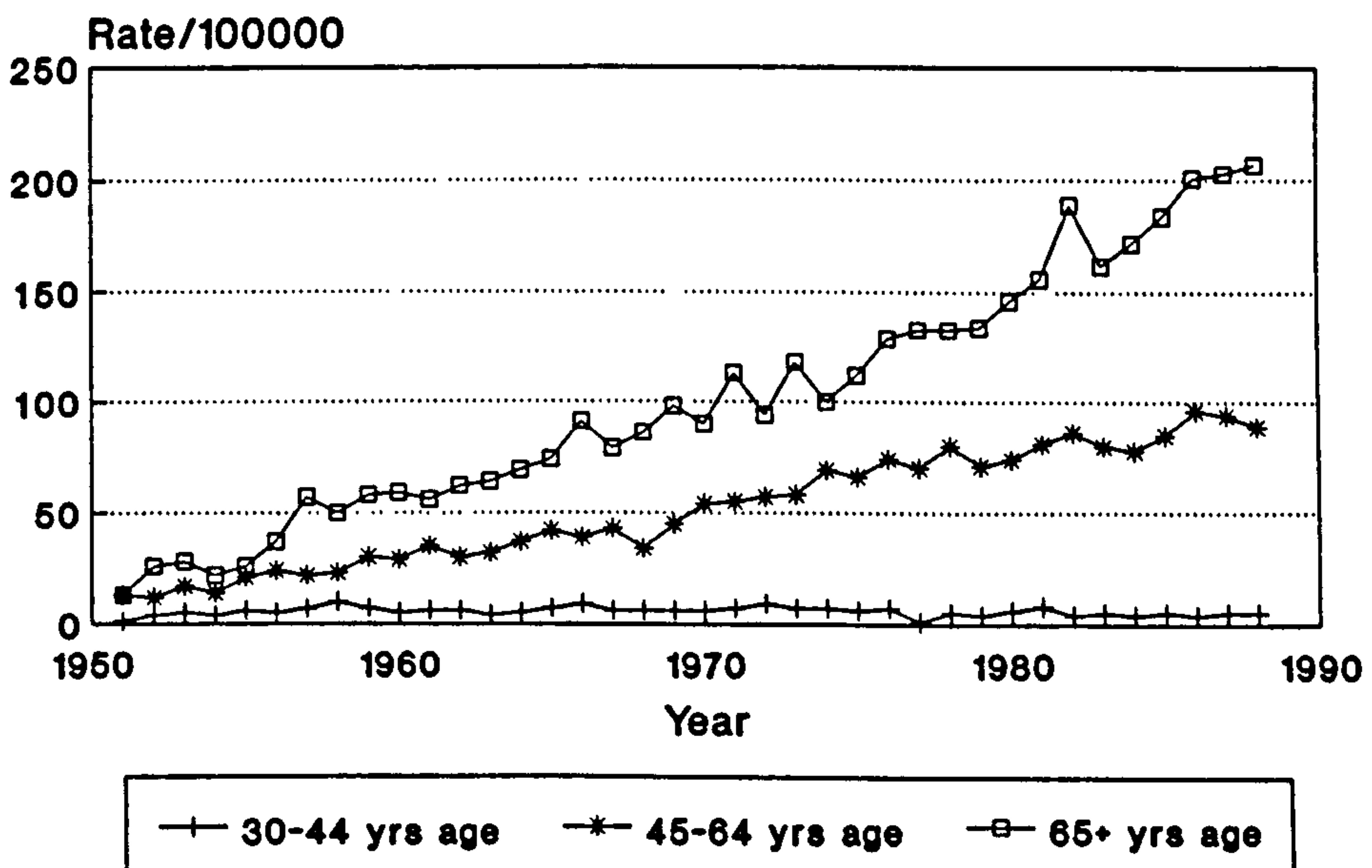
The 45-64 year age group (the middle aged group) follows more closely the pattern of the age-standardised rates. In male the broad-band age-specific rates increase from 140.0 per 100,000 in 1951 to a plateau of around 250 per 100,000 during the 1960s and early 1970s, and after 1975 decline steadily to 161.2 per 100,000 in 1988 (Figure 4.4). In females the broad-band age-specific incidence rates increase steadily from 12.9 per 100,000 in 1951 to 88.9 per 100,000 in 1988 (Figure 4.5).

**Figure 4.4 Broad-Band Age-Specific Lung Cancer Incidence Rates, Male**



Mersey Regional Cancer Registry

**Figure 4.5 Broad-Band Age-Specific Lung Cancer Incidence Rates, Female**



Mersey Regional Cancer Registry

In the 65 years and older age group (the old age group) the broad-band age-specific rates for males increase from

172.7 per 100,000 in 1951 to a plateau of approximately 750.0 per 100,000 in the late 1960s, after which period the rates remain relatively stable (Figure 4.4). For females the broad-band age-specific rates increase rapidly until 1957, from 13.1 per 100,000 in 1951 to 56.7 per 100,000 in 1957 (Figure 4.5). After 1957 the incidence rates increased exponentially at a rate of 2% per year, and by 1988 the incidence rate was over 200 per 100,000.

## 4.6 Age-Period Plots

### 4.6.1 Methods

The simplest, and most common, presentation of trends over time involves plotting the age-specific incidence rates against the central year for each quinquennium. The age-specific rates vary over several orders of magnitudes, therefore to aid in the interpretation of these plots, the logarithms of the rates are plotted against the logarithm of age. It has been shown that for many cancer sites the relationships between the log of the rates and the log of the age are linear (Juel 1983). Where these age-specific curves appear to be parallel any changes over the periods affect each age group in a similar manner. Any causative agent, therefore, affects all age groups in the same fashion with an effect on the



incidence, which is either immediate or at some fixed lag period.

The five year age-specific rates are presented for each quinquennium (period) in the log scale. For male lung cancer incidence rates are plotted for age groups starting at 30-34 years, and for female lung cancer incidence the youngest age group is 35-39 years.

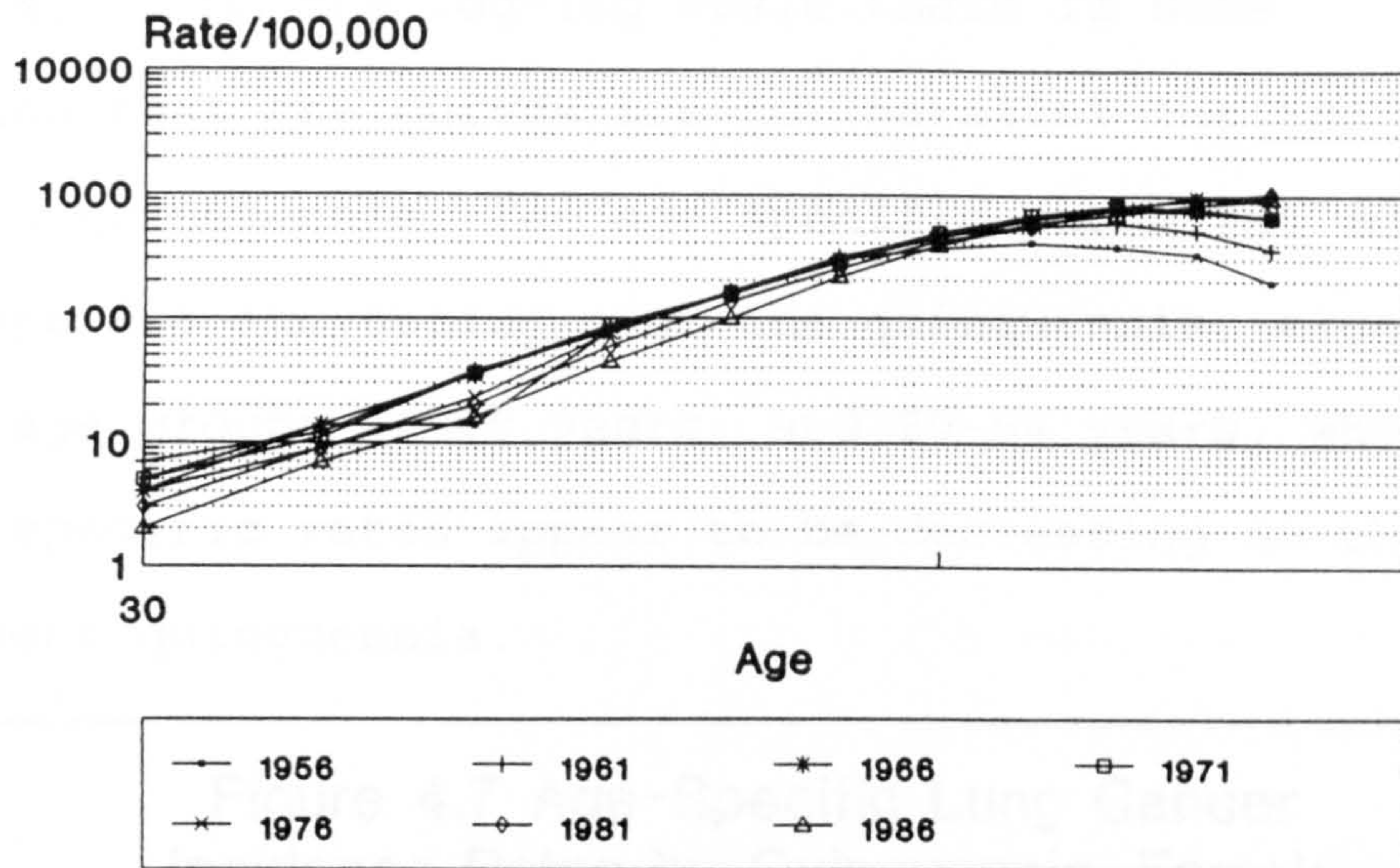
#### 4.6.2 Results

The five year age-specific lung cancer incidence rates for males begin to increase slowly from the age of 45 years, up to the age of 60 years (Figure 4.6). Over this age range there is little change in the age-specific rates over calendar time, except for the last two quinquennia (1981 and 1986) which are lower than those for the preceding quinquennia, and with that for 1986 lower than that for 1981.

In males over the age of 65 years, two marked patterns are seen in the age-specific rates (Figure 4.6). Before the quinquennia centred on 1976 the age-specific rates have their maximum in the 70-74 age group (604 per 100,000 for 1961 and 833 per 100,000 for 1971) except 1956, which has a maximum in the 65-69 age group of 408 per 100,000. For the three most recent quinquennia, the age-specific



**Figure 4.6 Age-Specific Lung Cancer Incidence Rates by Quinquennia, Male**



Quinquennia centred on census years  
Mersey Regional Cancer Registry

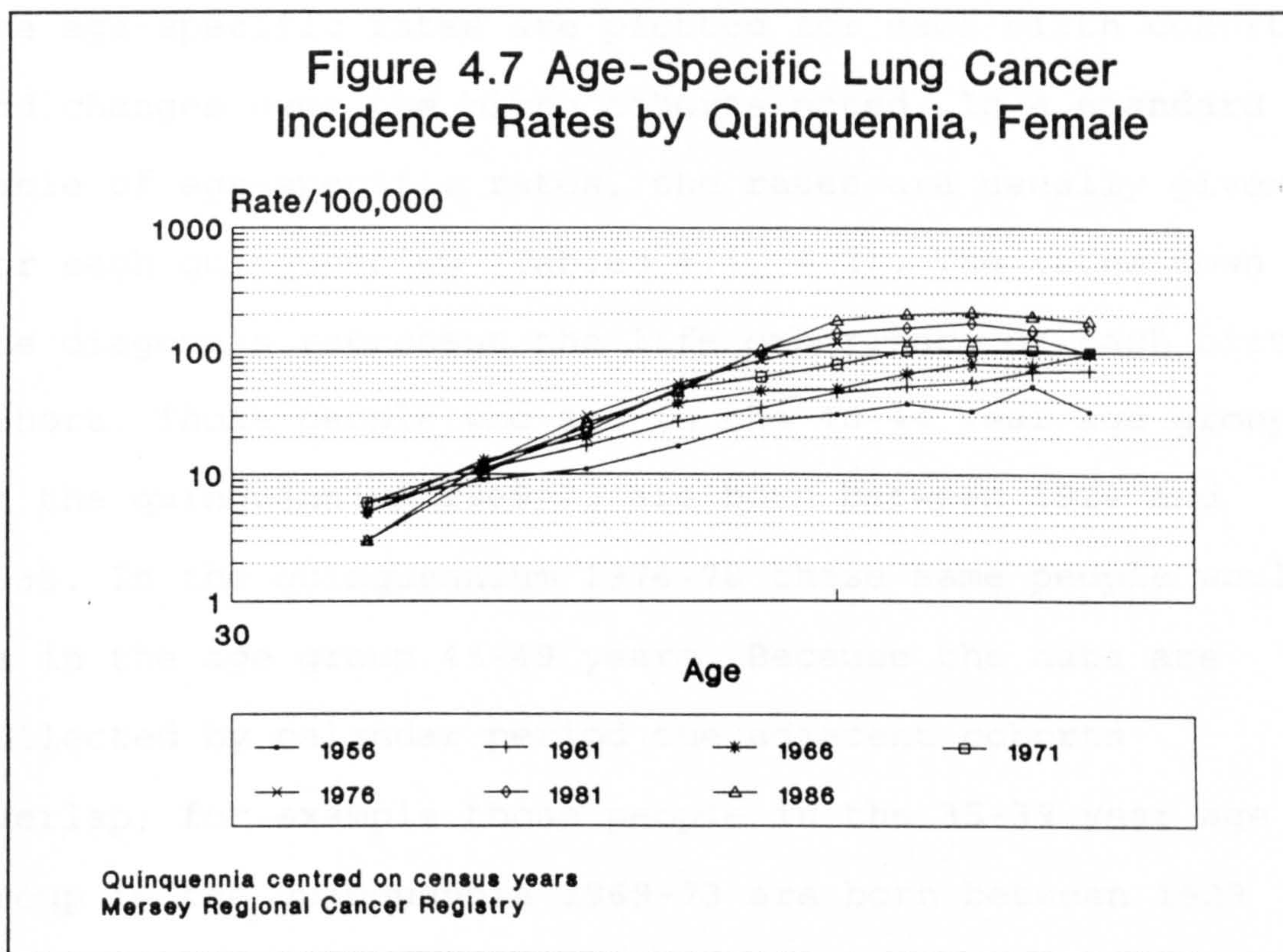
incidence rate increases steadily from the 55-59 year age group onwards, with no maximum apparent. For males over the age of 65 years the age-specific incidence rates show an increase over calendar time from 1956 to 1971, and then these age-specific rates begin to decrease over the next three quinquennia (Appendix IV(2)).

Examination of the log-log plot of age-specific incidence rates for males show parallel curve for the quinquennia up to the age of approximately 60. After the age of 60 the curves diverge in a similar pattern to that of the plot in the natural scales. This would indicate that the changes in male lung cancer incidence can not be explained only by changes over calendar period.

The five year age-specific incidence rates for female



lung cancer show an increase from the age group 35-39 years to the age group 65-69 years, for all quinquennia (Figure 4.7). On the log-log scale there is some indication that the curves are not parallel for the different quinquennia. The age-specific rates for most age groups are increasing over the quinquennia, except for the age groups 45-49 years, and 50-54 years, where the age-specific rates appear to be decreasing in the most recent quinquennia.



After the age of 65 years the rate of increase of the incidence rates over age slows. There is an indication that, for the more recent quinquennia, the rates reach a maximum at around the age of 70-74 years. The incidence rates, for those aged 65 years and over, increase over the time periods, with the rate for the 70-74 year age



group increasing from 57 per 100,000 in 1961 to 211 per 100,000 in 1986, a four-fold increase.

## 4.7 Age-Cohort Plots

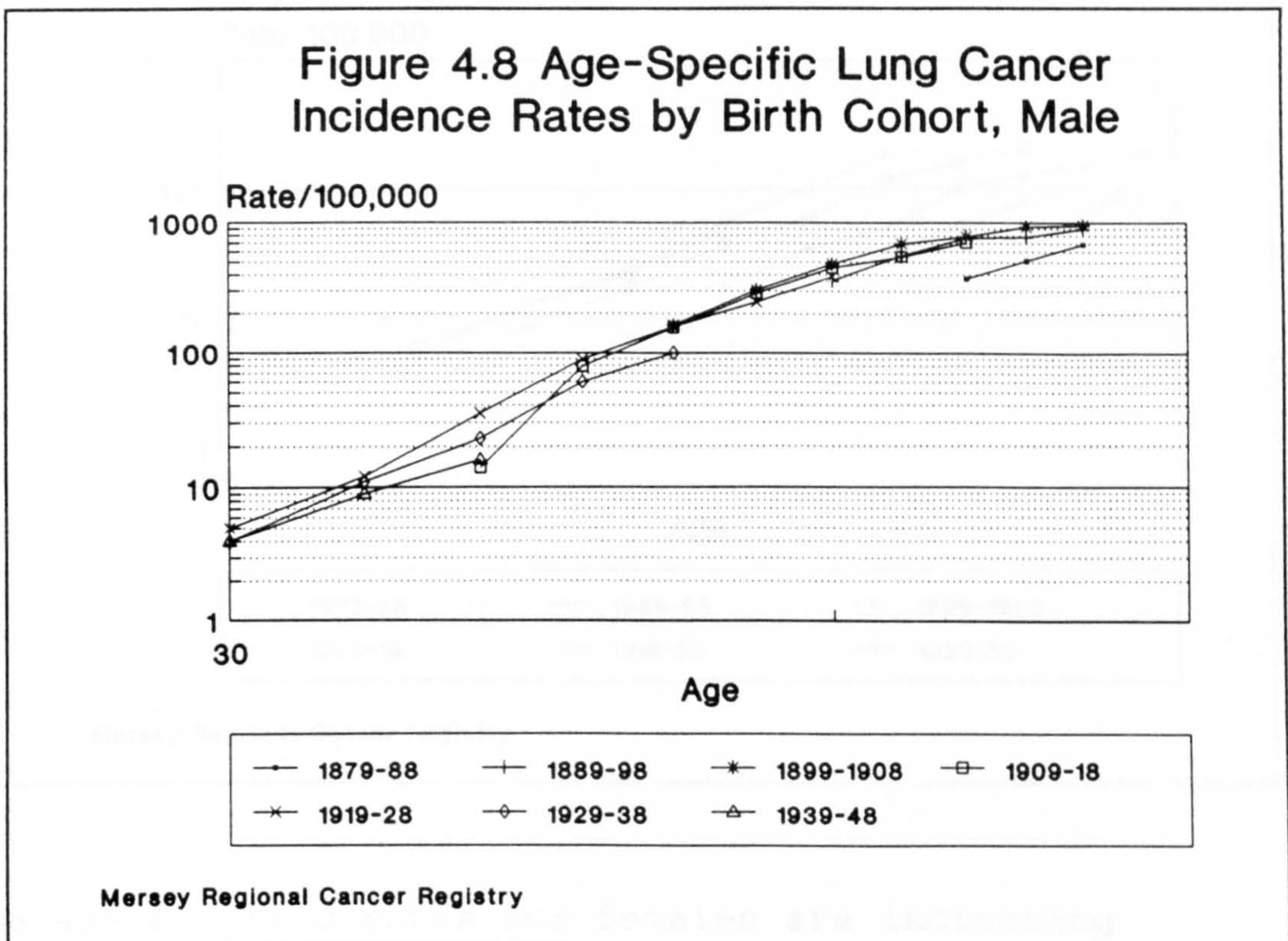
### 4.7.1 Methods

Another way to interpret the data is to examine the life time experiences of the different birth cohorts. For this the age-specific rates are plotted for each birth cohort and changes over the birth cohorts noted. In a standard table of age-specific rates, the rates are usually given for each quinquennium (Tables 4.1, 4.2). The rates down the diagonals represent the life experience of each birth cohort. Those people who are in the 40-44 year age group in the quinquennium 1969-73 are born between 1924 and 1933. In the quinquennium 1974-78 these same people would be in the age group 45-49 years. Because the data are collected by calendar period the adjacent cohorts overlap; for example those people in the 35-39 year age group in the quinquennia 1969-73 are born between 1929 and 1938. However, in most cases the rates down the diagonals give adequate estimates of cohort-specific rates. If the curves of logarithm of the rates versus the logarithm of age are approximately parallel then it can be assumed that any changes in risk behaviour influence the rates in each cohort equally throughout life.



The five year age-specific rates are plotted for every alternate five year birth cohort on both the natural and the log scales. Alternate cohorts are plotted for clarity. Little information is lost by not plotting all the cohort because the changes over cohorts tended to be smooth. For male lung cancer incidence rates are plotted for the age groups starting at 30-34, and for female lung cancer the starting age group is 35-39.

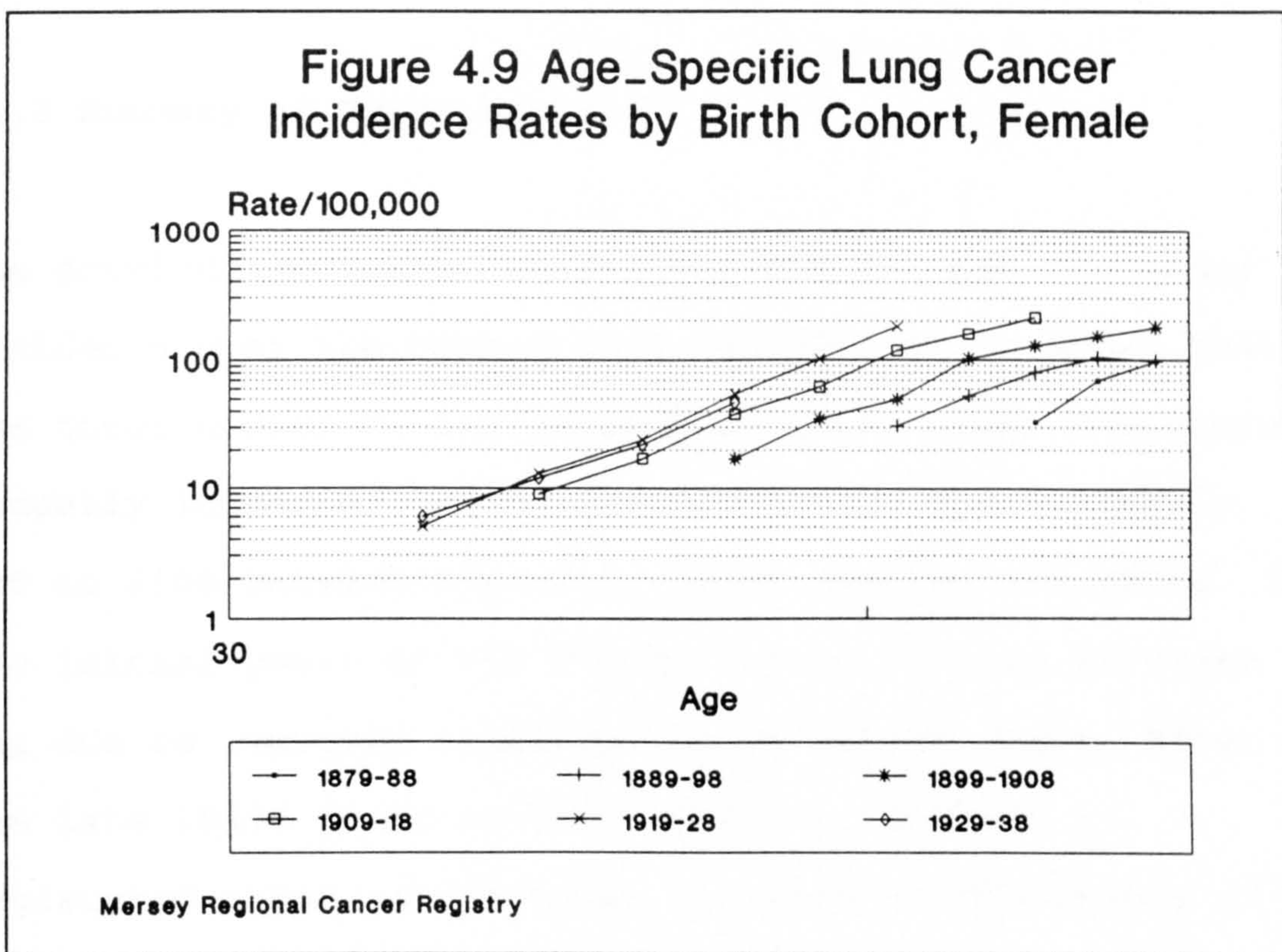
#### 4.7.2 Results



Cohort plots of male lung cancer incidence show a slight but consistent increase over the early cohorts (1879-88, 1889-98 and 1899-1908) followed by a consistent but even less marked decline over the more recent cohorts (Figure 4.8). The most striking difference is amongst the oldest



cohorts; 513 per 100,000 for the 75-80 age group in the cohort born between 1879-88 compared to 785 per 100,000 for the same age group for the cohort born between 1889-98. For the more recent cohorts no clear trend is discernible but this may be due to paucity of data; this cohort is still young and there have been few incident lung cancer cases.



The age-specific rates for females are increasing dramatically over the cohorts (Figure 4.9). The incidence rate in the 70-74 year age group has increased from 33 per 100,000 in the 1879-88 cohort to 211 per 100,000 in the 1909-18 cohort. The only exception to this is for the most recent cohort examined where there is an indication of a decrease. As with the males this may be due to



paucity of the data, because this cohort is not yet old enough to have entered the age range where the risk of lung cancer is high.

When the cohort patterns are examined in the log scale the age-specific curves appear approximately parallel. This indicates that the changes in female lung cancer incidence are probably due to factors affecting cohorts.

#### 4.8 Summary of Trend Analyses

The previous sections investigate the changes in cancer incidence over the period 1951 to 1988. It is shown that the total number of incident lung cancer cases registered annually in Mersey Region increased from approximately 500 to 2100 between the early 1950s and the mid 1970s. In the initial years of the registry part of this increase was due to improved registration of cancer cases. After the late 1950s it is assumed that the increase in registered cases is due to an increase in occurrence of lung cancer in the region. After the mid 1970s the total number of lung cancer cases has remained relatively constant; the decrease in the number of male lung cancer cases being compensated for by an corresponding increase in the number of female lung cancer cases.

In order to adjust for the changing population size and structure in Mersey Region five different forms of

incidence rates have been examined. These are the crude rates, the age-standardised rates, the broad-band age-specific rates , age-period plots and age-cohort plots.

The overall pattern for males is for these rates to increase in the 1960s and the 1970s, then to plateau until they began to decline in the 1980s. The only exception to this pattern is for the older males (65+ years) where the broad-band age-specific rates remained constant in the more recent years. The female lung cancer incidence rates increased over the whole period, for all age groups. The only exception to this pattern is for the most recent cohorts of females, where there may be an indication of declining incidence rates. However this must be interpreted with caution because these females are still young and only experiencing low lung cancer incidence.

#### **4.9 Projections**

Linear extrapolations of the trends in either the age-standardised incidence rates, or the broad-band age-specific incidence rates are accepted methods for projecting future lung cancer incidence (Teppo et al 1974, Marrett et al 1986). Projected incidence rates will be calculated by regressing the log of the appropriate rate against calendar year (Draper and Smith 1981).

In the calculation of the projection equation it is necessary to determine the range of years most appropriate for future extrapolation. Three sets of cancer incidence data are used for the projections. They are data from 1964-88, 1969-88 and 1974-88. The resulting equations are used for projections to the years 1990, 2000 and 2010. The methods used for the projections are now described.

#### **4.10 Projections for Age-Standardised Lung Cancer Incidence Rates**

##### **4.10.1 Methods**

Future cancer incidence rates for the year 1990, 2000 and 2010 are estimated by regressing the log of the age-standardised rate against calendar year using least squares regression (Draper and Smith 1981). The projected population for the year 2000 is used as the standard. The projected rates for the year 2010 are used to determine whether Mersey Region will achieve the targets set by the Health of the Nation for lung cancer (Department of Health 1991).

Although the 1981 population is used as the standard to investigate trends during the period of 1951 to 1988, it is necessary to use the projected population for the year 2000 if the projected number of incident cancer cases are



require for that year. The projected frequencies for the year 2000 are useful to determine service needs for that year.

The regression equation investigated is:

$$\ln(R^s_j) = \eta + \beta (y-1900)$$

where  $R^s_j$  is the age-standardised incidence rate for period j

$\eta$  is the constant of regression

$\beta$  is the gradient of the regression equation, and

y is the year of diagnosis

95% confidence limits are determined by:

$$\ln(R^s_j) \pm t_{n-2,0.05} s(\ln(R^s_j))$$

where

$s(\ln(R^s_j))$  is the standard error for the log of the projected age-standardised rate given by

$$s(\ln(R^s_j)) = \text{sqrt}[(s^2_{R.Y}) \times \{1 + (1/n) + ((y - \bar{y})^2 / \Sigma (y - \bar{y})^2)\}]$$

and

$s^2_{R.Y}$  is the residual mean square error for the regression

n is the number of data points used in the regression and

$\bar{y}$  is the mean of y

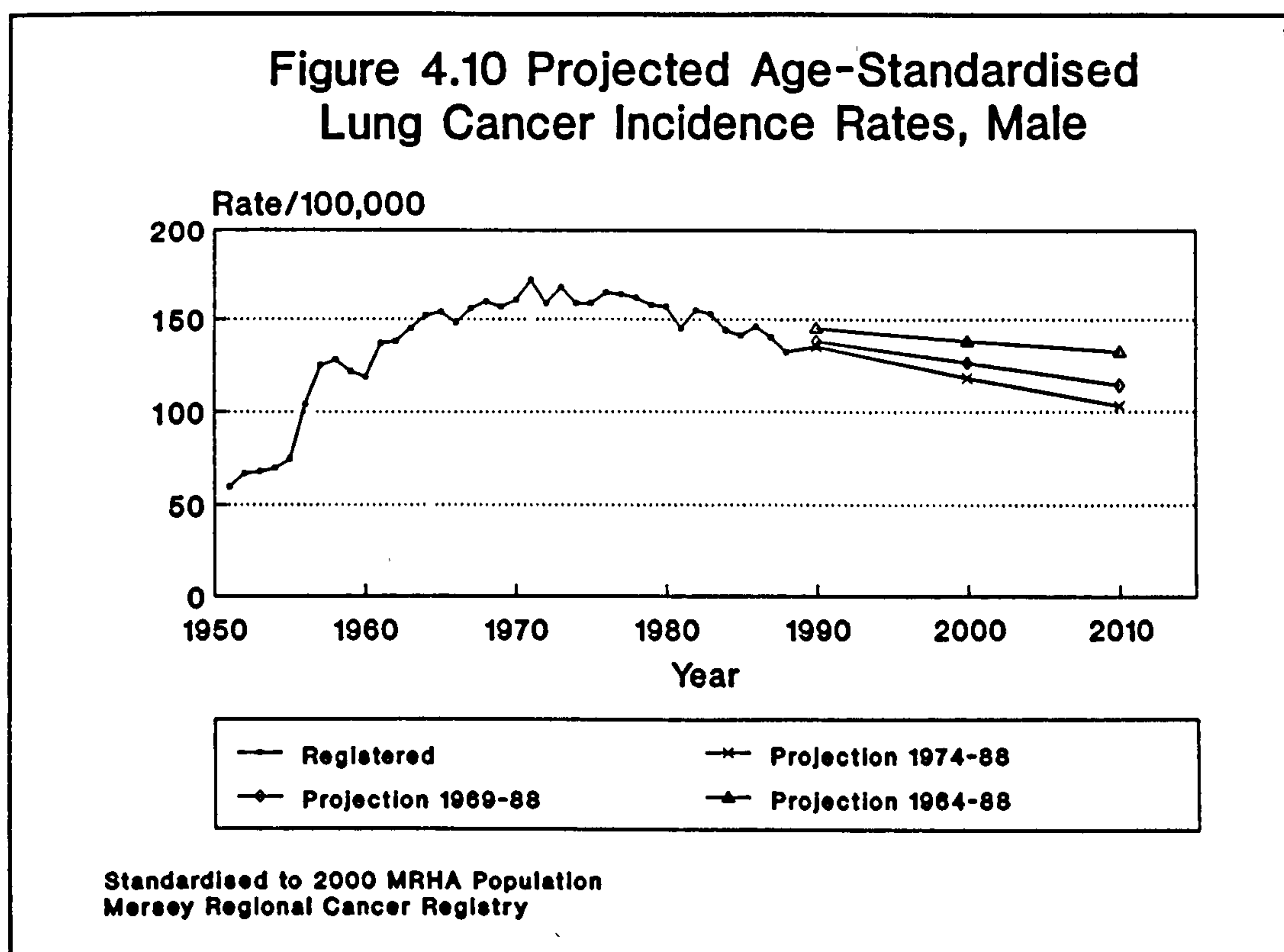
The number of cases expected in the year 2000 are determined by multiplying the projected age-standardised

rate by the projected population.

Three sets of incidence data are used for these projections; 1964-88, 1969-88 and 1974-88.

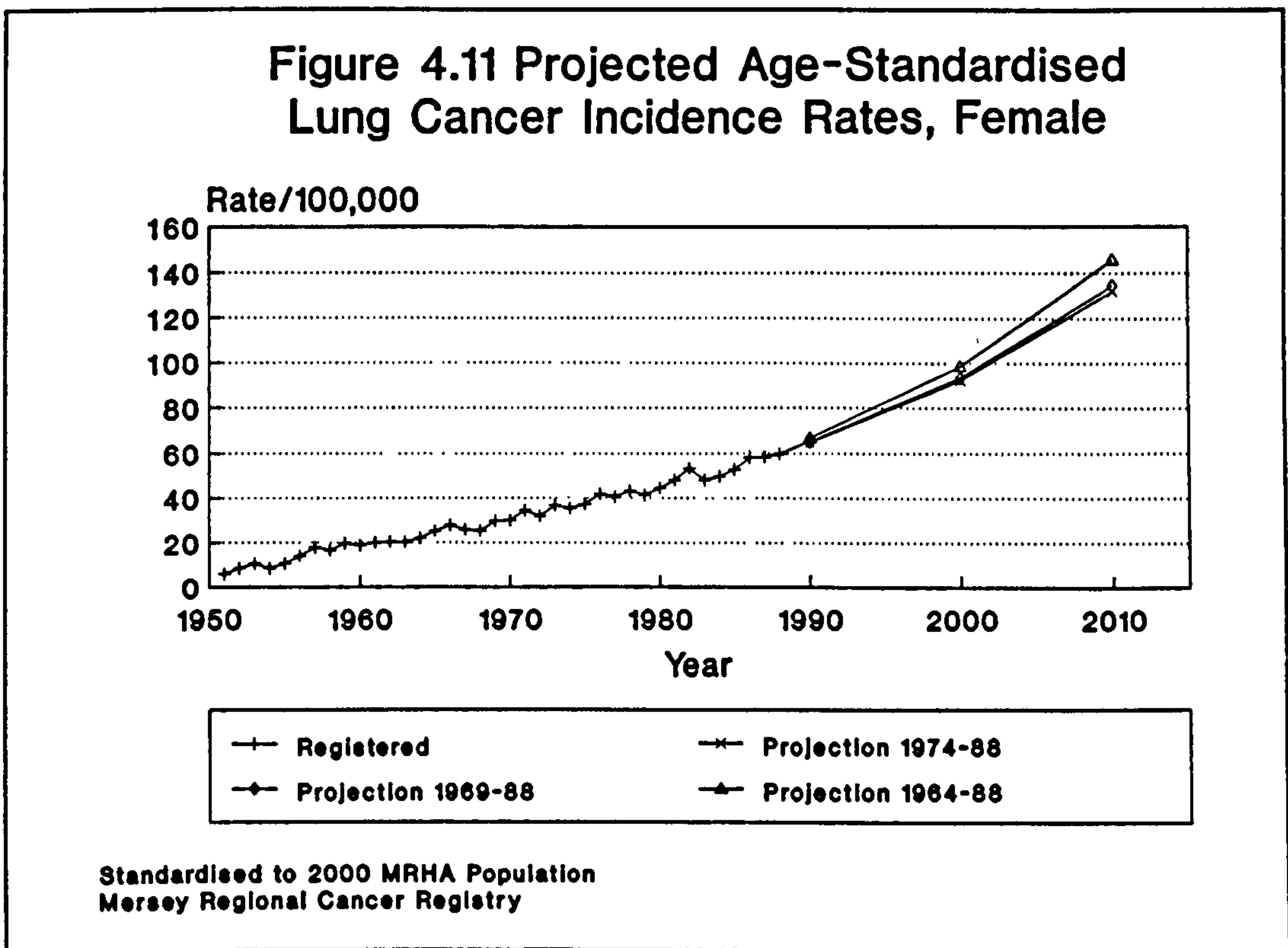
#### 4.10.2 Results

The use of the projected population for the year 2000, instead of the 1981 population, does not affect the overall trends in the age-standardised rates (Figures 4.10, 4.11).



For males the projections based on the 1974-88 data give the closest agreement with the age-standardised rate obtained from the number of cases actually registered in 1990 (134.6 cases per 100,000 for both estimates). The

projected age-standardised incidence rates based on the longer data sets 1969-88 and 1964-88 are higher than those obtained from the 1974-88 data set (Figure 4.10). This reflects the non-linear trend in the male rates. The problems of non-linearity are also reflected in the increasing width of the confidence intervals as more years are included in the regression (Appendix IV(3)).



The projections, using the three different time periods, are more consistent for females than for males. This is due to the consistent increase in lung cancer incidence amongst females over the last 40 years. The projected age-standardised rates for 1990 are all similar to age-standardised rate of 65.5 cases per 100,000 obtained from those cases actually registered in 1990 (Figure 4.11). As with the males, the projections based on the longer data



sets are slightly higher than those based on the 1974-88 data set. This may reflect a slight lessening of the upward trend in recent years.

The projected age-standardised rates for males decline over the period 1990 to 2010 for all three data sets, indicating an overall decline in lung cancer incidence from the early 1960s. For females the age-standardised rates all increase from 1990 to 2010. In 2010 the age-standardised rates for females are projected to be higher than those for males (132 per 100,000 and 103 per 100,000 for females and males respectively, using data from 1974-88)

Due to the presence of non-linearity amongst male lung cancer incidence prior to the early 1970s only the projections based on the 1974-88 data set will be discussed in relation to the Health of the Nation and the service needs. These projections gives age-standardised rates of 135 male cases per 100,000 and 65 female cases per 100,000 for the year 1990; 118 male cases per 100,000 and 92 female cases per 100,000 for the year 2000; and 103 male cases per 100,000 and 132 female cases per 100,000 for the year 2010. This would mean 1367 cases of male lung cancer and 1123 cases of female lung cancer in the year 2000.

#### 4.10.3 Health of the Nation Targets

If this recent trend in male lung cancer incidence is to continue, the target to reduce lung cancer mortality, and hence incidence, by at least 30% by 2010 would almost be achieved; the projected reduction being 28.5% since 1990. However, for females the age-standardised incidence rates are projected to double over the same time period. These projections are based on the total population and not the population for those under 75 years, which is the target population for Health of the Nation. Even though the incidence of lung cancer is high in those aged over 75 years, this age group only accounts for a relatively small proportion of the population (2.5%). Thus, there should be little difference between the percentage change in age-standardised rates for the two populations.

#### 4.10.4 Service Needs

The projections for service needs using age-standardised rates is only approximate, because the age-specific rates are not available. Using the overall rates for treatment (Section 3.5, Appendix III(3)), these projections would indicate that in the year 2000, 134 males and 110 females would require surgery, 131 males and 108 females would require chemotherapy and 247 males and 203 females would require radiotherapy. Between 1983 and 1987, 150 males and 60 females underwent surgery on average each year,

135 males and 72 females received chemotherapy and 272 males and 118 females received radiotherapy. Thus although there would be a decrease in service needs for males, this decrease is more than compensated by the projected increase in the service needs for females.

#### **4.11 Projections for Broad-Band Age-Specific Lung Cancer Incidence Rates**

The second method that is described for projecting lung cancer incidence is the extrapolation of the trends in the broad-band age specific rates.

##### **4.11.1 Methods**

The number of lung cancer cases that occur each year in the broad-band age groups can be assumed to be Poisson distributed (Clayton and Schiffler 1987a). The model relating the broad-band age-specific rates to the year of diagnosis is fitted by maximum likelihood techniques (McCullagh and Nelder 1989). GLIM software is used (Baker and Nelder 1978).



This relationship can be modelled by:

$$\ln(R_{ij}) = \ln(\mu_{ij}) - \ln(N_{ij}) = \eta_i + \beta_i (y-1900)$$

where  $R_{ij}$  is the age-specific incidence rate for age group  $i$ , and year  $j$ ,

$\mu_{ij}$  is the expected age specific frequency for age group  $i$ , and year  $j$ ,

$N_{ij}$  is the age-specific person years at risk for age group  $i$ , and year  $j$ ,

$\eta_i$  is the constant for age group  $i$ , and

$\beta_i$  is the coefficient of regression for age group  $i$

This model assumes that the observations are independent, which may not be the case here.

In matrix notation this is given by:

$$\ln(R) = (\ln(\mu) - \ln(N)) = X\beta$$

where

$R$  is the vector of age-specific incidence rates  $R_i$

$\mu$  is the vector of age-specific frequencies  $\mu_i$

$N$  is the vector of age specific person years at risk  $N_i$

$X$  is the design matrix, and

$\beta$  is the vector of coefficients

Person years at risk are estimated by the population size ( $N_{ij}$ ) multiplied by 1 year and included in the model with a coefficient of unity. The asymptotic confidence

intervals are calculated, using an assumption of normality, by the method proposed by Hakulinen and Dyba (1994) where:

$$I_{ij} \pm z_{0.05} s(I_{ij})$$

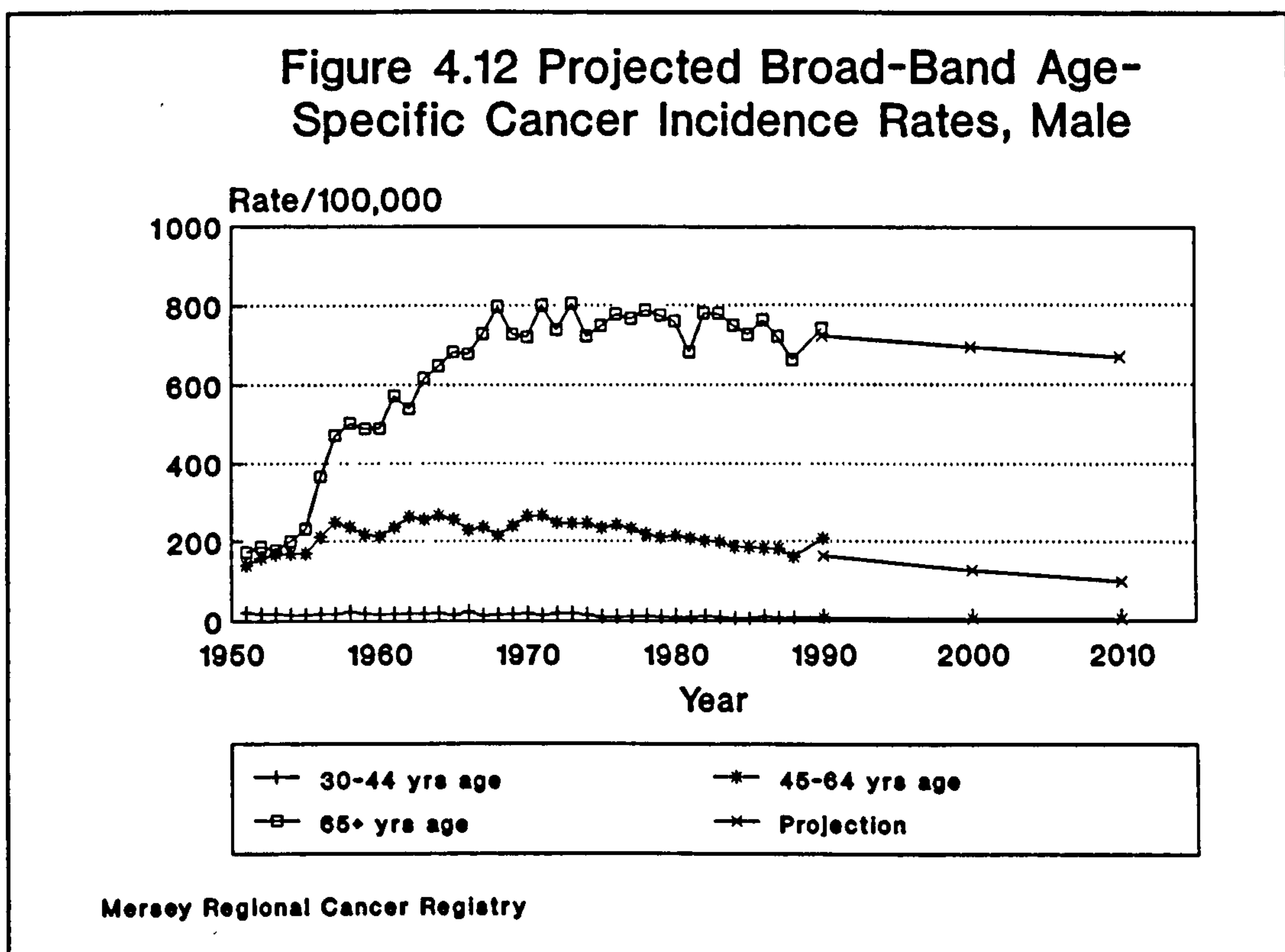
where  $s(I_{ij}) = \text{sqrt}(\mu + X(X'W^{(x)}X)^{-1}X')$

and  $W^{(x)}$  is the matrix of weights used in the last iteration. In the case of the Poisson distribution these are the estimates of the fitted values.

As with the age-standardised rates data from three data sets are used, 1964-88, 1969-88, and 1974-88. For each data set the above models are fitted for each age group. The total cancer load for each sex are estimated from the sum of the age-specific estimates. The age-specific projections for 1990, for each data set, are compared with the age-specific frequencies of registered lung cancer cases. The projections for the year 2010 are compared with those for 1990, in order to determine whether the targets for the Health of the Nation would be reached. Service needs for the year 2000 are determined from the age-specific projections, assuming that the treatment regimens would be the same as those in 1983-87 (Williams et al 1993).

#### 4.11.2 Results

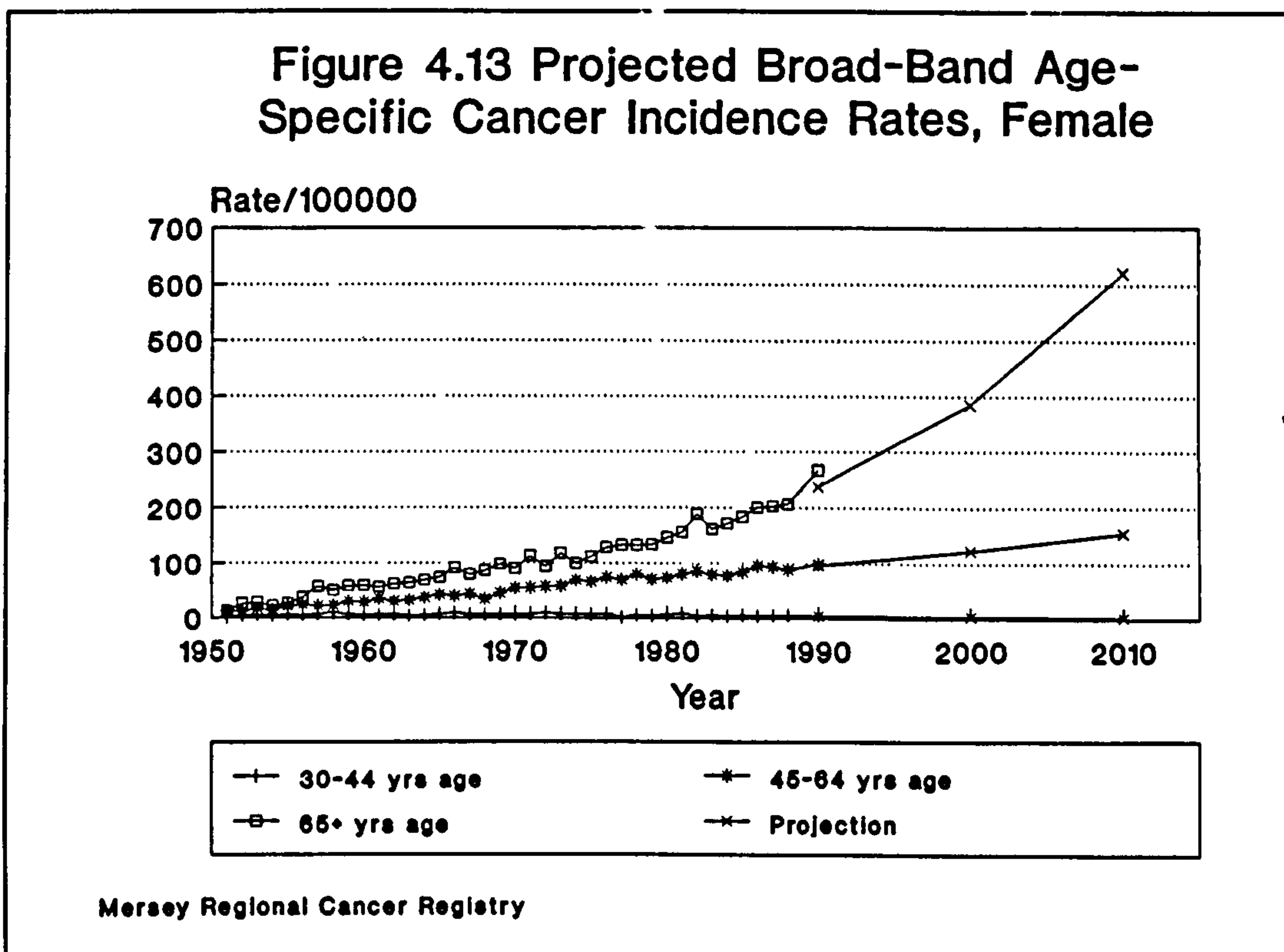
The projections based on the broad-band age-specific rates for males reflect the decline in the rates for all aged groups from the mid 1970s onwards (Figure 4.12). The only exception to this pattern are for those projections for the older age group based on data from 1964 to 1988 which indicate a rising incidence rate (Appendix IV(4a)). This discrepancy in the older age group is due to the non-linearity of lung cancer incidence during the period 1964 to 1988.



The projected numbers for females are more consistent than those for males, and show an increasing trend for all age groups (Figure 4.13). As with the age-standardised projections there is little difference among



the three data sets (Appendix IV(4b)). In the year 2000, the projections for the total number of female lung cancer varied between 1214 and 1277. These are somewhat higher than the projected numbers based on the age-standardised incidence rates of 1123 incident cases in the year 2000.



The total number of male lung cancer cases projected for 1990, using the 1974-88 data set, agrees most closely with the number registered for that year. 1629 lung cancer cases were registered and the projected figure for males is 1479 (Appendix IV(4a)). The difference is due to an under estimation in the middle age group. For females the corresponding figures are 874 and 802 cases (Appendix IV(4b)). However, for males the projected incidence rates are over estimated in the 65+ age group and under

estimated in the 45-64 year age group. For females, the situation is reversed with over estimation in the 45-64 year age group, and underestimation in the 65+ year age group. Examination of the broad-band age-specific incidence rates in recent years shows considerable variation around the general trends, and the rates for 1990 appear to be high in comparison to past trends. Thus variations between the projected rates and the observed rates could be expected for any single year.

#### **4.11.3 Health of the Nation Targets**

The projections based on the 1974-88 data sets indicate that the targets for the reduction of lung cancer incidence by 30% for the young and middle aged males would be achieved. For the older males the reduction in incidence rates are only 7.5%. For females, the targets for Health of the Nation will not be met. For models based on the 1974-88 data sets, the projected increases are 78%, 68%, and 204% for the 30-44 year, 45-64 year and 65+ year age group respectively.

#### **4.11.4 Service Needs**

The projections based on the broad-band age-specific rates allow service needs to be estimated for the year 2000. The projections based on the 1974-88 data set are used for these estimates. It is assumed that the

proportion of cases that are treated with surgery and/or chemotherapy and/or radiotherapy remain the same as those for 1983-87. The number of cases projected to receive surgery in 2000 would be 130 males and 109 females, the number of cases receiving chemotherapy would be 118 males and 131 females, and 246 males and 223 females would undergo radiotherapy. These figures would indicate an overall increase in service requirements, particularly in relation to radiotherapy. The different percentage increases are due to the increasing age of the lung cancer cases, and the differentials in treatment profiles for the different age groups.

#### **4.12 Summary of Projections**

In recent years lung cancer incidence in MRHA has decreased in males and increased in females (Section 4.9). Projected lung cancer incidence rates are calculated based on the extrapolation of trends in age-standardised incidence rates and broad-band age-specific incidence rates.

These projections based on the age-standardised rates indicate that male lung cancer incidence is likely to decline by almost 30% between 1990 and 2010. The same is true for projections based on broad-band age-specific incidence rates. Therefore the Health of the Nation targets will be achieved, if the same trends are



maintained. Obviously, with projected increasing lung cancer incidence, females will not meet the Health of the Nation target for lung cancer incidence

If the more recent trends in male lung cancer incidence are maintained, it is projected that service needs for male lung cancer patients will decrease. If the projections are based on age-standardised rates service needs for surgery will decrease by 10%, for chemotherapy by 3% and for radiotherapy by 9%. If they are based on broad-band age-specific rates service needs will decrease by approximately 13% for surgery and chemotherapy, and by approximately 10% for radiotherapy. Service needs for female lung cancer are projected to increase. Projections based on age-standardised rates indicate that service needs will increase by 83% for surgery, 50% for chemotherapy and 72% for radiotherapy. If the projections are based on broad-band age-specific rates service needs will increase by over 80% for all forms of therapy. For both methods of projection total service needs will increase by approximately 20% for each type of therapy.

All of the above projection are based on the assumption that the changes in rates are due to calendar period effects. Examination of the age-period plots and the age-cohort plots indicate that although there appear to be changes over period, there are also changes that are cohort specific, and that the more recent birth cohorts

in both males and females are experiencing lower rates of lung cancer. Thus the above projections may be over estimating the future lung cancer burden.

Chapter 5

**TRENDS IN SMOKING BEHAVIOUR**



The previous chapter investigates the trends in lung cancer incidence during the period 1951 to 1988. The trends in lung cancer incidence are thought to be related to previous trends in smoking behaviours of the population.

This chapter explores the trends in smoking variables that reflect the changing patterns of smoking behaviour.

In particular, this chapter will:

- 
- describe the trends in the number of manufactured cigarettes smoked per person per year for each sex separately, and for the age groups 25-34 years and 35-59 years for each sex separately
  - describe the trends in the percentage of the population aged over 16 who smoke cigarettes for each sex separately, and for the age groups 25-34 years and 35-59 years for each sex separately
  - describe the trends in the number of manufactured cigarettes smoked per smoker per week for each sex separately, and for the age groups 25-34 years and 35-59 years for each sex separately
  - describe the trends in the percentage of smokers who smoke plain cigarettes for each sex separately
  - describe the trend in the sales adjusted average tar yield per cigarette
-

The data on these smoking variables must be population based and also available for an adequate duration.

### 5.1 Methods

As stated in Section 3.4 all the data on smoking are abstracted from UK Smoking Statistics (Wald and Nicolaides-Bouman 1991).

The number of cigarettes smoked per person, the percentage of the population who smoke and the number of cigarettes smoked per smoker are plotted against year, for each sex separately.

Smoking is usually taken up in the late teens and early twenties, and the number of cigarettes smoked per person or per smoker increases in the first few years (Wald and Nicolaides-Bouman, 1991). Therefore, data on the age group 25-34 years would represent the experience of the young to middle aged population. As the population ages, some smokers will change their smoking habits, such as quitting smoking or reducing the number of cigarettes smoked, for health or other reasons. Therefore the age group 35-59 years would represent the experience of the older population.

The age groups for the data on age-specific consumption of cigarettes per person change over time (Section 3.4,

Table 3.5). Data for the age group 25-34 years are available only for the years 1975-87, and for the age group 35-59 years only the years 1949-55. The numbers of cigarettes smoked per person for the age groups 25-34 years and 35-59 years are estimated using the weighted averages of the age-specific rates given and the population figures for Great Britain as the weights. For example:

$$CP_{25-34,1960} = \frac{N_{25-29,1960} \times CP_{25-29,1960} + N_{30-34,1960} \times CP_{30-34,1960}}{N_{25-29,1960} + N_{30-34,1960}}$$

where  $CP_{25-34,1960}$  is the average number of cigarettes smoked per person per year in age group 25-34 years in 1960

and  $N_{25-34,1960}$  is the population of Great Britain in age group 25-34 years in 1960

Data on the percentage of the population smoking is available for the 25-34 year age group for the years 1948-87 (Section 3.4) The percentages for the age group 34-59 are only available up to 1975. For the years after 1975 the percentages for this age group are estimated using the weighted average of the age-specific percentages given (age groups 35-49 years and 50-64 years) with the population figures for Great Britain as the weights. This is similar to the method used for cigarettes per person.



Information on the number of cigarettes smoked per smoker is available for the same age groups as the number of cigarettes smoked per person (Section 3.4). The numbers of cigarettes smoked per smoker for the age groups 25-34 years and 35-59 years are using the weighted averages of the age-specific consumption figures with the number of smokers in Great Britain as the weights. For example:

$$CS_{25-34,1960} = \frac{N_{25-29,1960}P_{25-29,1960}CS_{25-29,1960} + N_{30-34,1960}P_{30-34,1960}CS_{30-34,1960}}{N_{25-29,1960}P_{25-29,1960} + N_{30-34,1960}P_{30-34,1960}}$$

where  $CS_{25-34,1960}$  is the number of cigarettes smoked per smoker in age group 25-34 years in 1960

$N_{25-34,1960}$  is the population for year 25-34 years in 1960

and  $P_{25-34,1960}$  is the percentage of people smoking in age group 25-34 years in 1960

Data on the percentage of smokers who smoke plain cigarettes are available for both males and females (Section 3.4) These data are available for 1958, 1961, 1963, 1965, 1968 and all the years between 1971 and 1987. Information on the total annual sales of filter and plain cigarettes is available for the years 1905 to 1987 (Wald and Nicolaides-Bouman 1991, Table 1.2). The relationship between the percentage of smokers who smoke plain cigarettes and percentage of total annual sales of

cigarettes that are plain is investigated, using linear regression. The fit is found to be excellent ( $R^2=0.99$  for both males and females). For the years where there is no data, the percentage of male and female smokers who smoke plain cigarettes are estimated from the regression equations. The resulting data on the percentage of the population who smoke plain cigarettes is plotted against year for each sex separately.

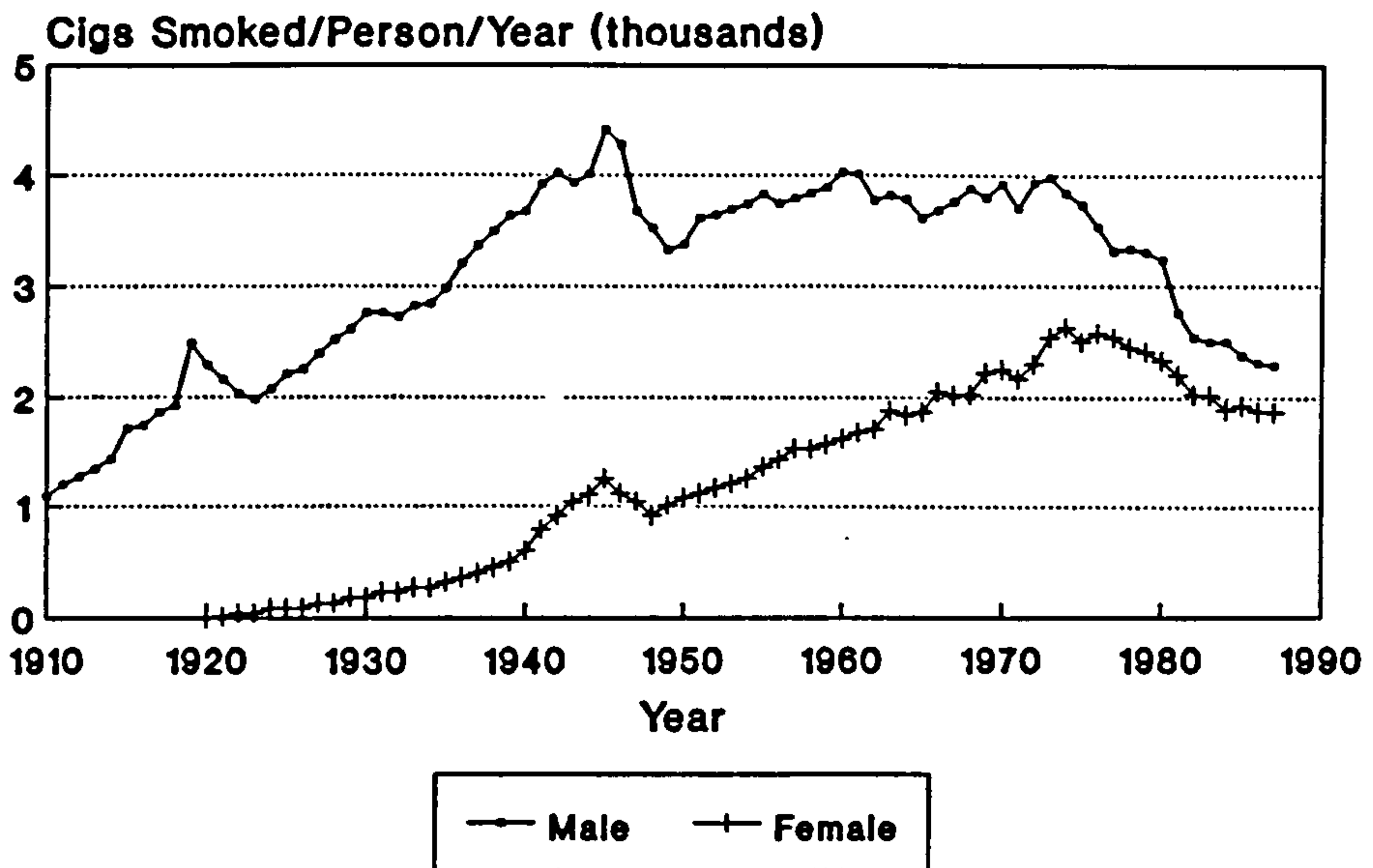
Data on the average tar content of cigarettes are available as annual estimates for the years 1970 to 1988, and for the years 1934 to 1969 as averages for seven year periods eg 1934-40 (Section 3.4). Yearly estimates are calculated using linear interpolation between the central points of the intervals. No adjustment was made for differences between the sexes because such data is only available from 1978.

## **5.2 Results**

### **5.2.1 Trends in Manufactured Cigarettes Smoked per Person per Year**

The number of cigarettes per person reflects the population consumption of cigarettes. Changes in this consumption are affected by two factors, the percentage of the population who smoke and the average number of cigarettes smoked per smoker.

**Figure 5.1 Manufactured Cigarettes Smoked per Person per Year, by Sex, UK**



TAC

The number of cigarettes smoked per male per year increases steadily from 1910 to 1960, rising from just over 1,000 cigarettes per male per year (or approximately 20 cigarettes per week) in 1910 to just over 4,000 cigarette per male per year (or approximately 80 cigarettes per week) in 1960 (Figure 5.1) During the war years, 1914 to 1919, and 1939 to 1945, there are addition peaks. The maximum amount smoked was 4420 cigarettes per male per year (or 85 cigarettes per week) in 1945. After both World Wars the total consumption fell over approximately a five year period to resume the previous trend. The consumption was fairly stable from 1960 to 1975, with a slight decrease in the early 1960s, followed by a slight rise in the late 1960s. After 1975 the consumption fell quite rapidly to around 2,300 cigarettes

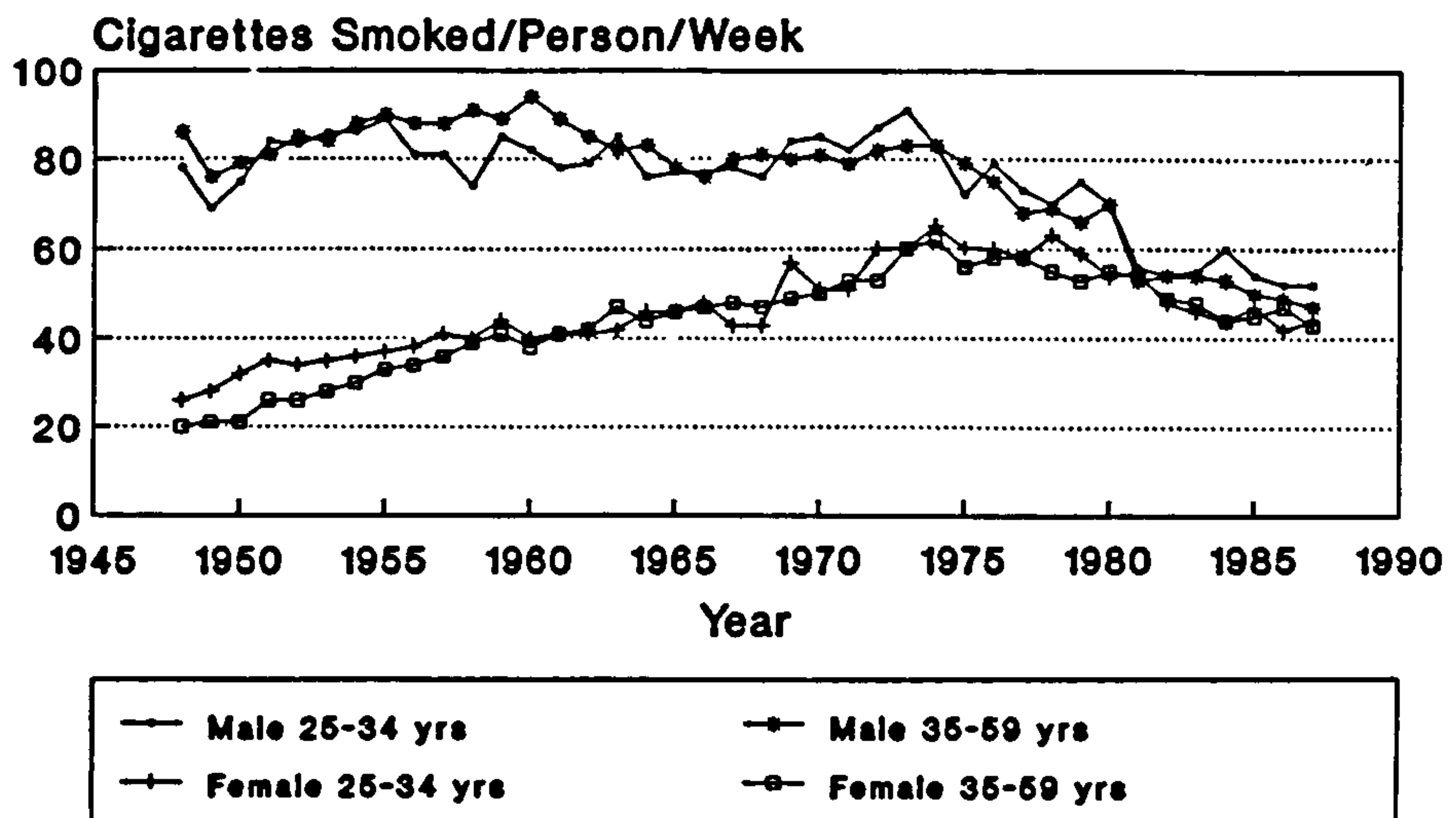


per male per year (or approximately 45 cigarettes per week) in 1987.

Data on the quantity smoked by females is available from 1920. Prior to that time smoking was predominantly a male activity. The average number of manufactured cigarettes smoked per female increased steadily from no cigarettes per female per year in 1920 to over 2,500 per female per year (or approximately 50 cigarettes per week) in 1974, except for a local peak during World War II (Figure 5.1). Between 1974 and 1978 there was little change in consumption. After 1978 the number of cigarettes smoked per female per year decreased to 1870 (or 35 cigarettes per week) in 1987. Females have always smoked less than males, but the difference between the sexes is decreasing.

There is little difference in the number of cigarettes smoked per person between the two age groups examined, 25-34 years and 35-59 years (Figure 5.2) As with the total number of cigarettes smoked per person females smoke less than males, although the difference between sexes is decreasing. Males in the older age group (35-59 years) smoked slightly more cigarettes than the younger age group (25-34 years) until the mid to late 1960s, after which the younger age group smoked more. The number of cigarettes smoked per male in the older age group increased from around 80 cigarettes per week in the early

**Figure 5.2 Manufactured Cigarettes Smoked per Person per Week, by Sex and Age, GB Sales Adjusted**



TAC

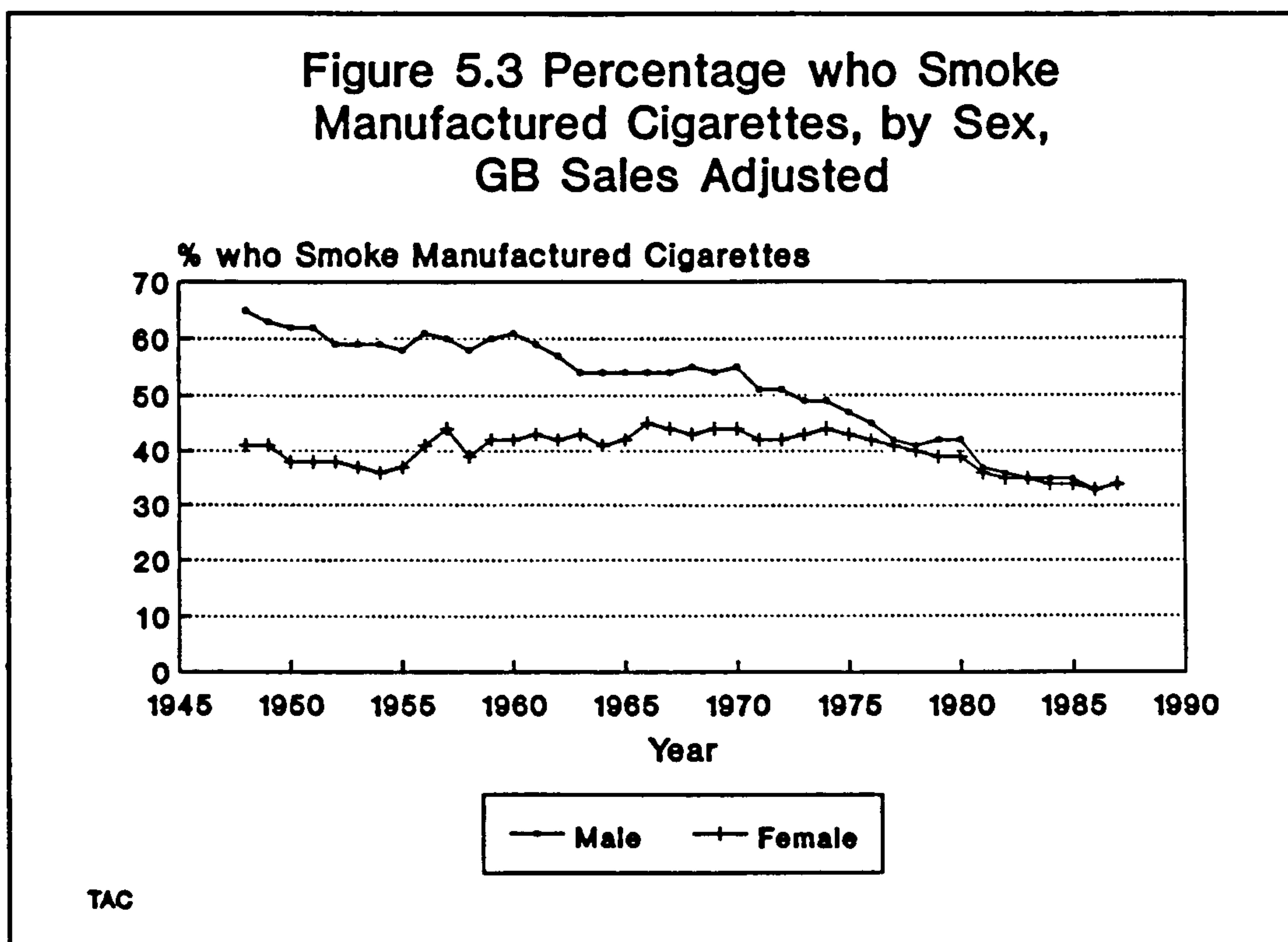
1950s to around 90 cigarettes per week in the early 1960s. The consumption in this age group then decreased to around 80 cigarettes per week by 1965, and then it remained stable until around 1975. After 1975 the amount smoked decreased to less than 50 cigarettes per week in 1987.

The pattern in the younger age group showed considerable variation around a constant level of approximately 80 cigarettes per male per week, up until 1975. After 1975 the consumption began to decline. By 1985 the consumption was just over 50 cigarettes per week.

The females consume fewer cigarettes than the males (Figure 5.2). However, the consumption rose steadily from around 30 cigarettes per week for the younger age group

and 20 cigarettes per week for the older age group in the late 1940s to around 60 cigarettes per week for both age groups in the early 1970s. The quantity smoked for the younger age group remained at this level until 1979 after which the consumption decreased to around 45 cigarettes per week by the late 1980s. For the older age group consumption started to decrease around 1977 and fell to around 45 cigarettes per week by the late 1980s.

### 5.2.2 Trends in the Percentage of People who Smoke Manufactured Cigarettes



The percentage of males who smoked declined from approximately 65% in 1948 to approximately 55% in 1970 (Figure 5.3). The rate of decline then increased rapidly, falling to less than 35% in 1986. These figures do not



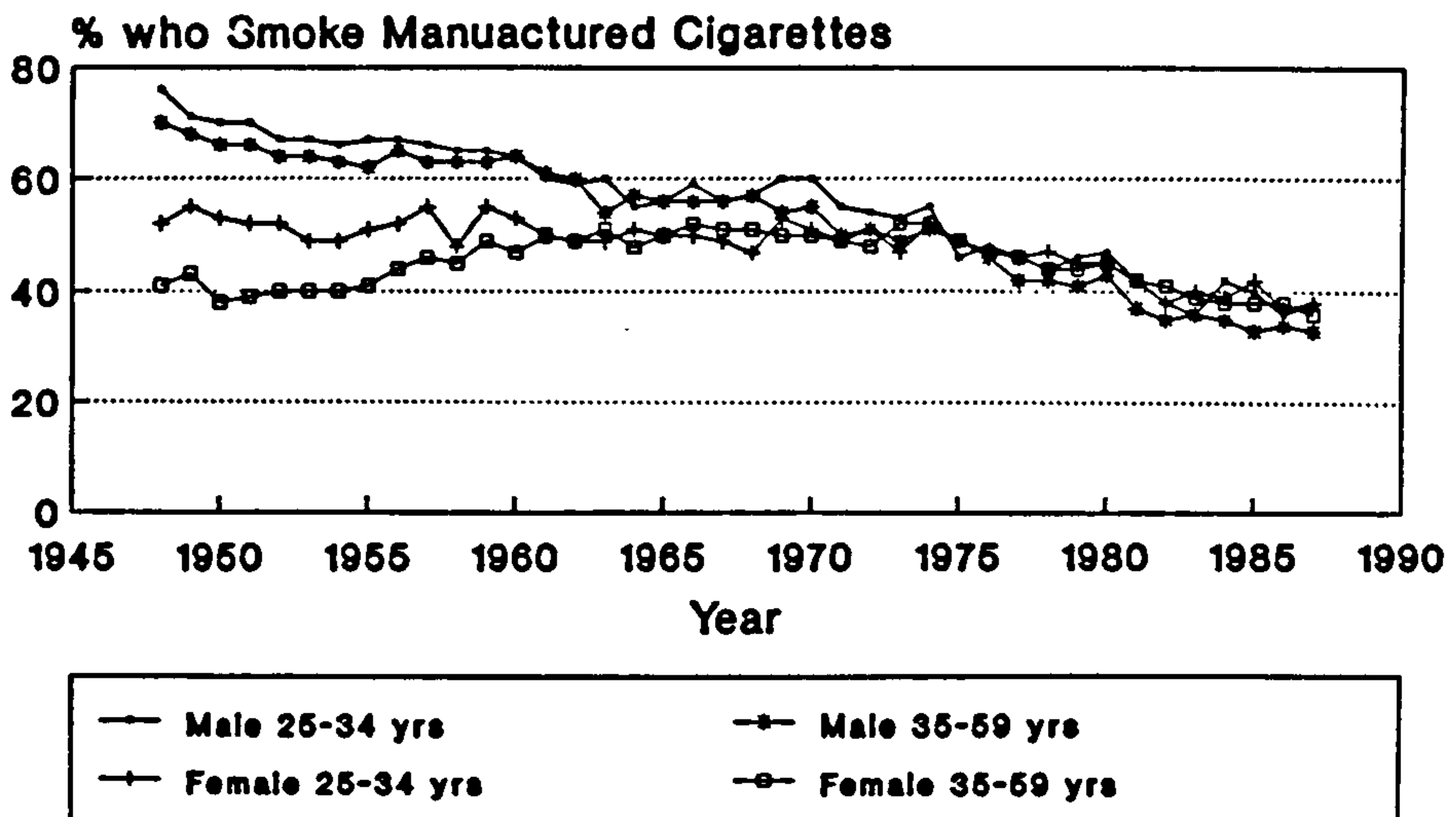
include those men who only smoke hand rolled cigarettes. From 1957 onwards the percentage who smoke hand rolled cigarettes has remained quite stable at 0.9% of the population. These smokers have been excluded from the analysis since data are not available for many of the characteristics of these smokers.

In 1948, over 40% of women smoked (Figure 5.3). This percentage dropped over the next five years to just over 35%, but subsequently increased to over 40% by 1957. The percentage of females who smoked remained relatively stable until the early 1970's. Since then there has been a steady decline in the percentage of women who smoke to 34% in 1987. Thus the same percentage of males and females are smoking. As with the males the statistics are for manufactured cigarettes only.

A higher percentage of males in the 25-34 year age group smoke than in the 35-59 year age group (Figure 5.4). This may be an effect of the healthy smoker phenomenon.

Individuals may quit smoking because of ill health. The percentage of the population who smoke has fallen steadily from the late 1940s, from around 70% in the late 1940s to around 38% for the younger age group and around 33% for the older age group in the late 1980s.

**Figure 5.4 Percentage who Smoke Manufactured Cigarettes, by Sex and Age GB Sales Adjusted**

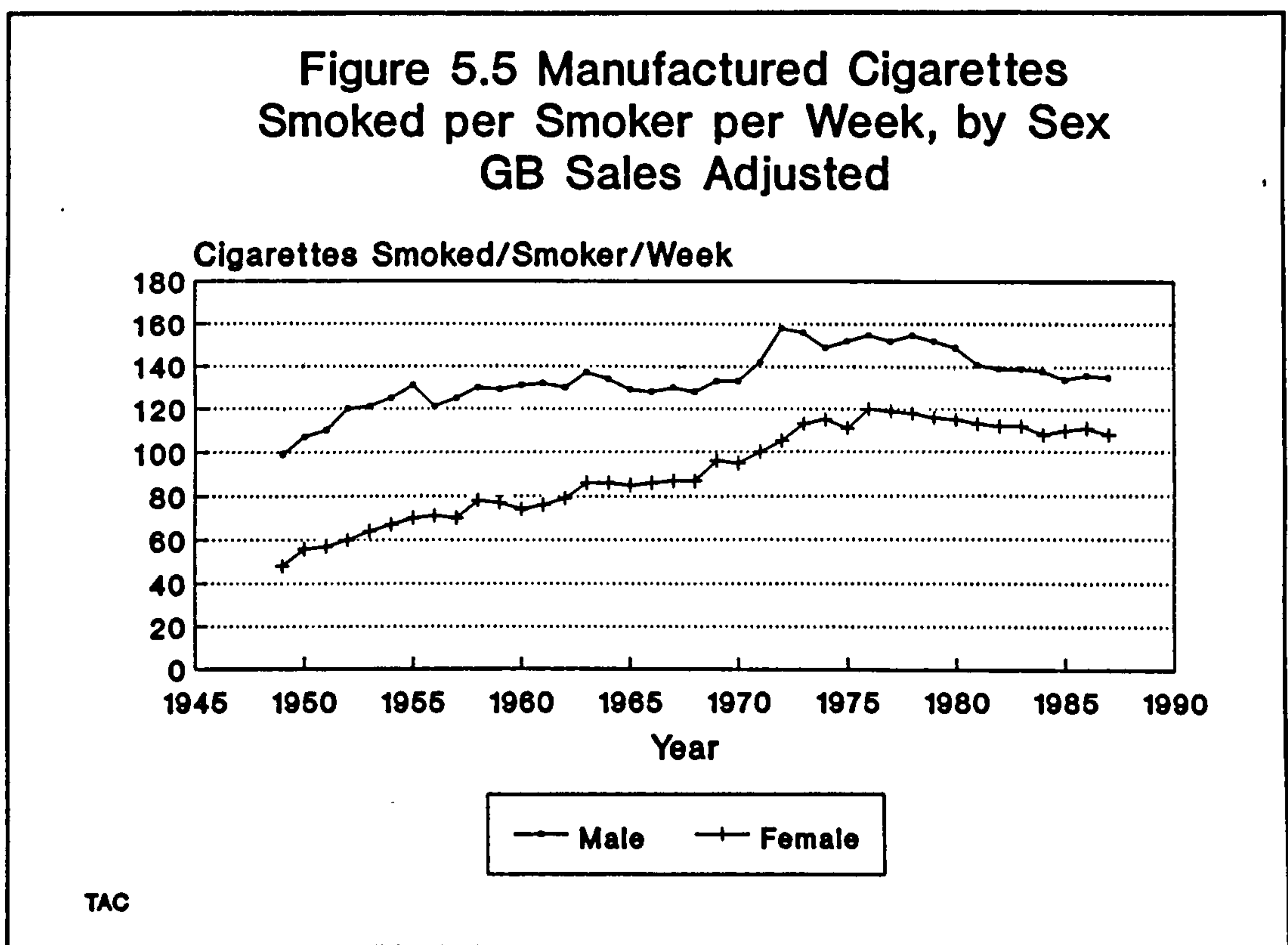


TAC

The percentage of females in the 25-34 year age group who smoked remained fairly constant at around 50% from 1948 to 1975 (Figure 5.4). After 1975 the percentage of females who smoked in this age group declined steadily to just under 40% by the late 1980s. The percentage females who smoked in the 35-59 year age group remained constant at around 40% until 1955, after which the percentage increased to around 50% by 1960. This is the same percentage as for the younger age group for that period, and reflects that those women who were in the younger age group in the 1950s would be in the older age group by the 1960s. After 1960 the trends for the two age groups are essentially the same, and by the late 1980s less than 40% of the females in both age groups were smoking.

### 5.2.3 Trends in the Number of Cigarettes Smoked per Smoker per Week

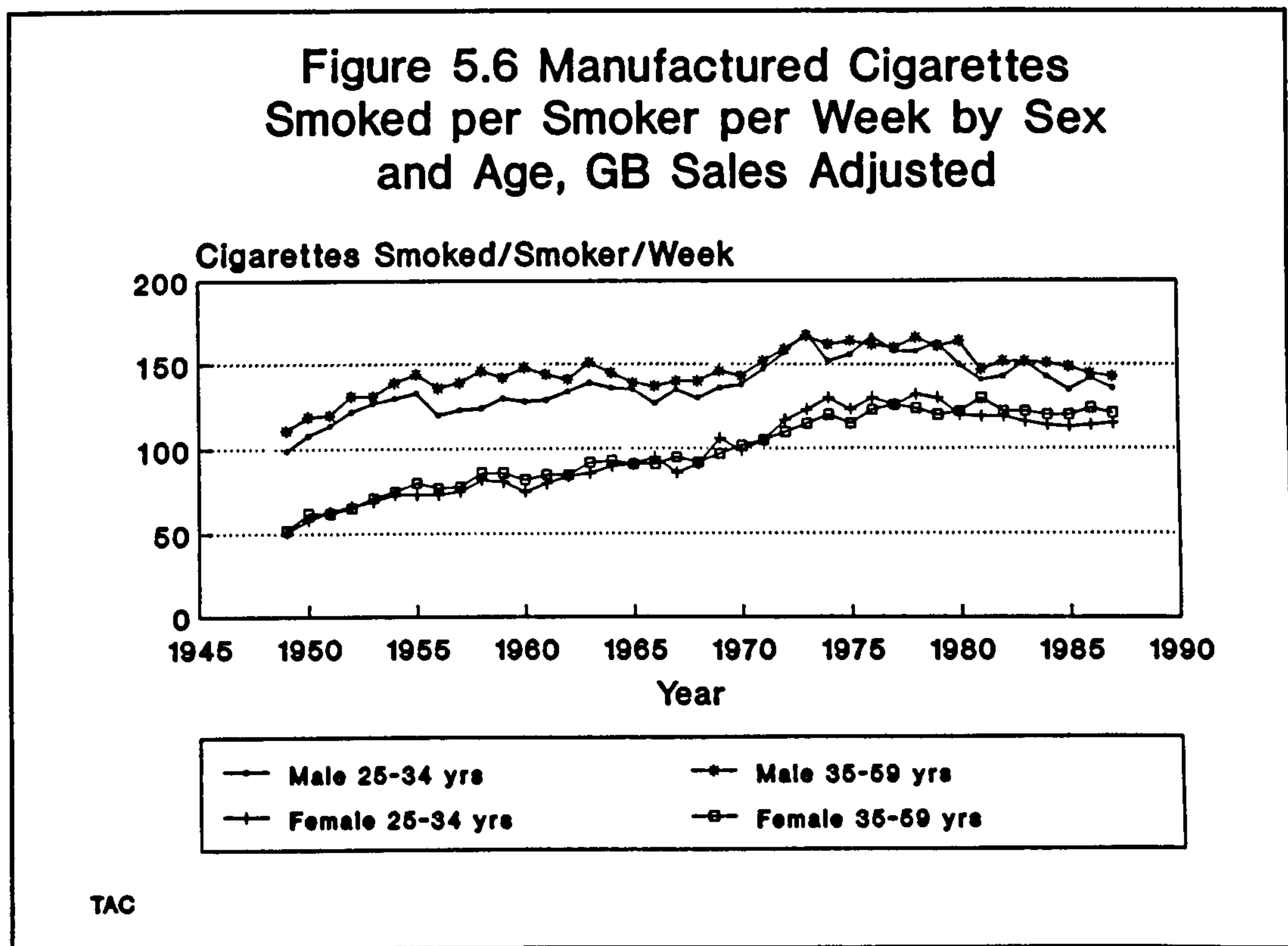
The number of cigarettes smoked per male smoker per week increased steadily from 1948 (approximately 100 cigarettes) to a maximum in the early 1970s (approximately 160 cigarettes) (Figure 5.5). Consumption of cigarettes began to decline in the early 1980s and fell to 135 cigarettes per week in 1987.



The pattern observed for females is very similar to that seen in males (Figure 5.5). However, the consumption levels are consistently lower. Females smoked approximately 50 cigarettes per week in 1949, 120 cigarettes per week in the late 1970s and just over 100 cigarettes per week in the late 1980s.



Males in the older age group smoked slightly more cigarettes per smoker than those in the younger age group for the whole period of interest (Figure 5.6). The consumption rose from around 110 cigarettes per week for the younger age group and 120 cigarettes per week for the older age group in 1950 to around 130 and 140 cigarettes per week respectively by 1955. After 1955 the consumption remained fairly constant until 1970 when consumption started to increase again to around 160 and 165 cigarettes per week respectively for the age groups by 1975. This consumption was maintained until 1980 when it dropped to around 140 and 150 cigarettes per week in 1987 for the age groups respectively.



Female smokers have consumed fewer cigarettes than their male counterparts, and there appear to be little

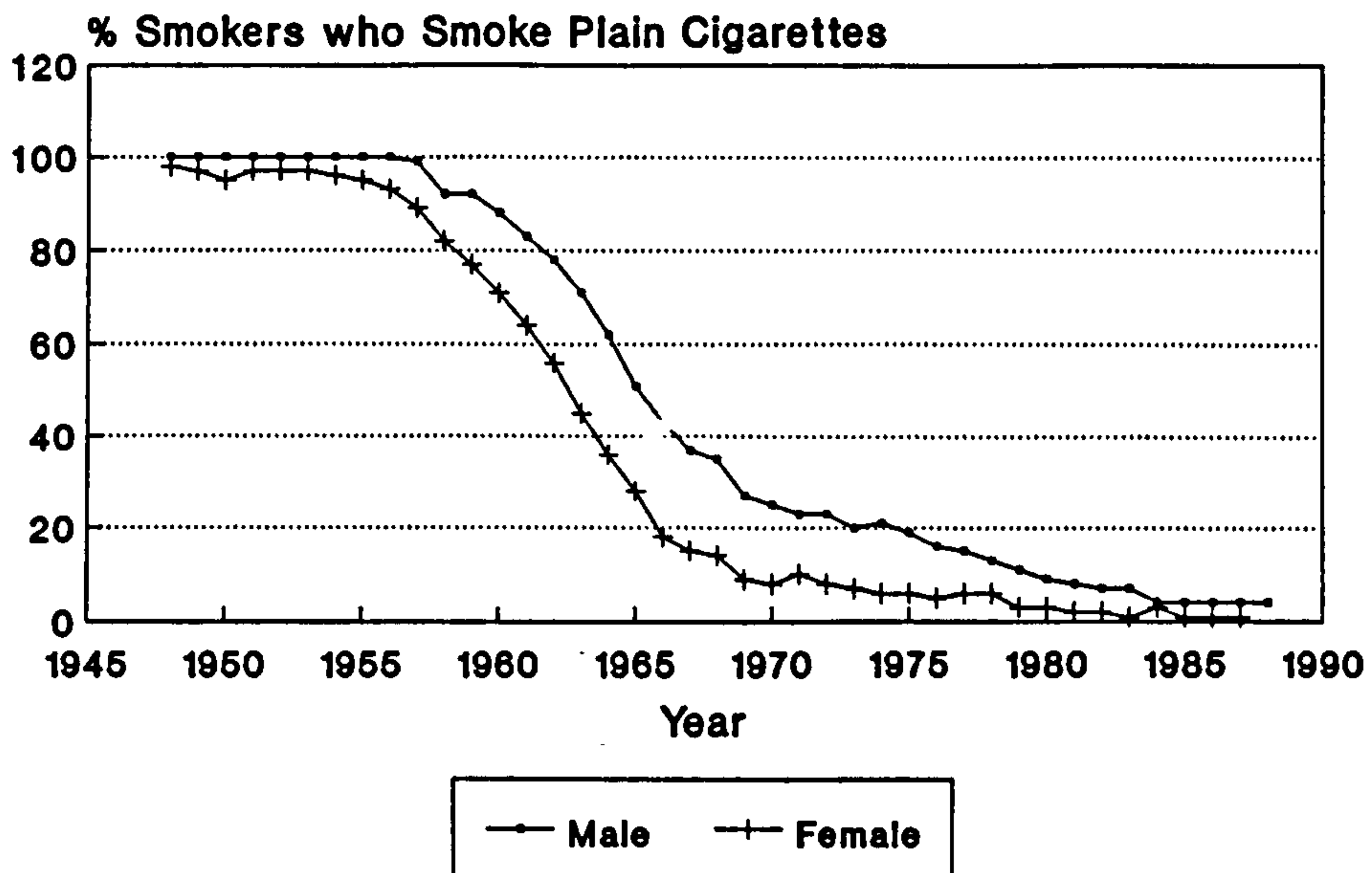
difference between the two age groups, 25-34 years, and 35-59 years. Consumption has risen steadily from just over 50 cigarettes per week in the late 1940s to around 125 cigarettes per week by 1975. After this time the consumption plateaued, although there may be evidence of a slight decline to around 120 in the more recent years.

#### **5.2.4 Trends in the Percentage of Smokers who Smoke Plain Cigarettes**

In the early part of the century only non filter cigarettes were available. Filter cigarettes were introduced in the 1930's but only became popular in the 1960's. Data are available from 1958 onward on the percentage of smokers who smoke plain cigarettes for each sex.

In 1958, 92% of the male smokers smoked plain cigarettes and 82% of females smokers (Figure 5.7). These percentages declined steadily until 1970 when 25% of male smokers and 10% of females smokers smoked plain cigarettes. After 1970 the rate of decrease slowed and over the next 15 years the percentage decreased to 4% for males and 1% for females. The percentages of male and female smokers who smoke plain cigarettes have remained constant since that time.

**Figure 5.7 Percentage of Smokers who Smoke Plain Cigarettes, by Sex, GB**



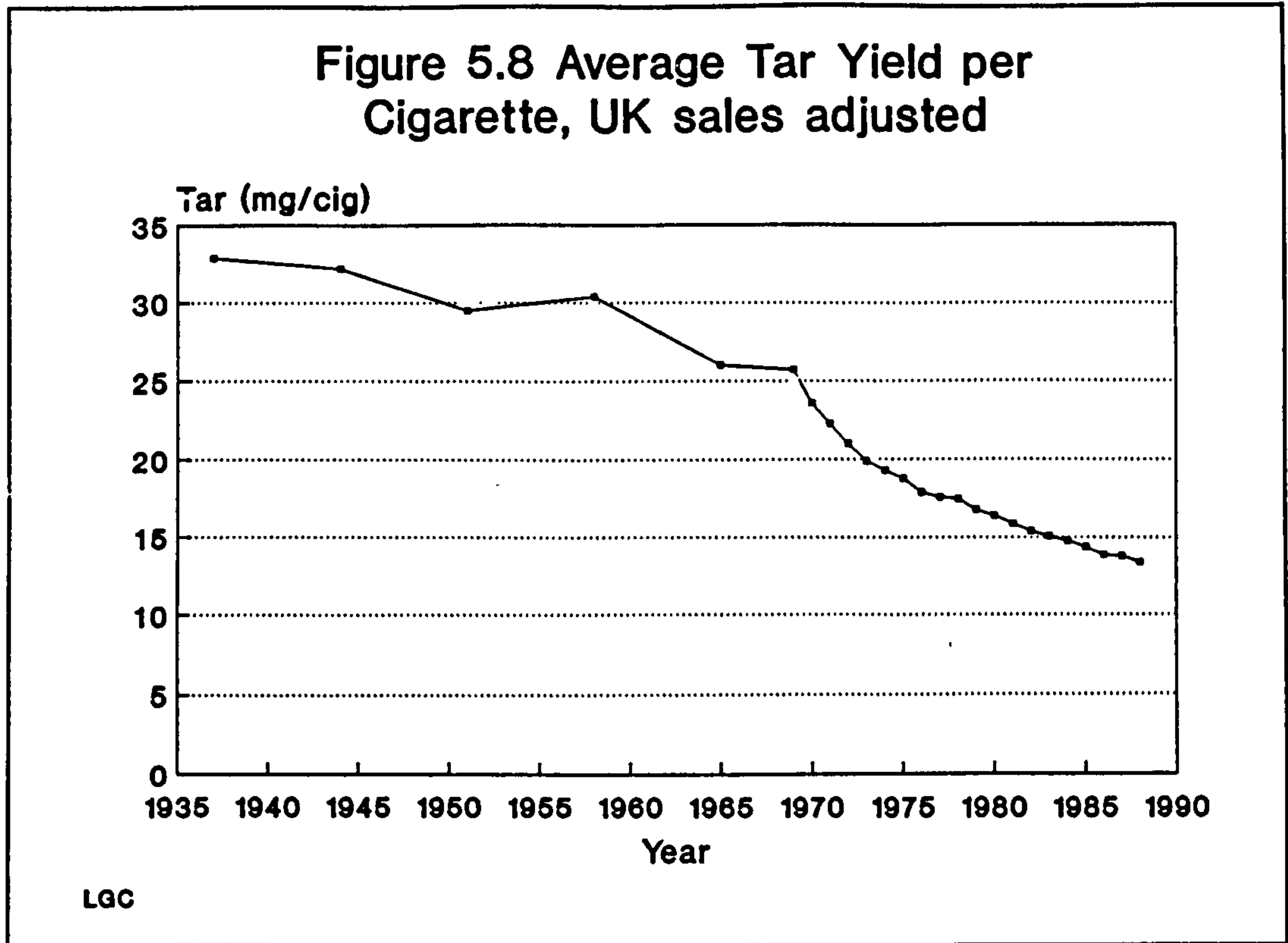
TAC

#### 5.2.5 Trends in Sales Adjusted Average Tar Content of Cigarettes

The average sales weighted tar content of cigarettes has fallen since the 1930s, from 32.9 mg per cigarette to 25.7 mg per cigarette in the 1969 (Figure 5.8). After the 1969 the rate of decrease became more rapid and by 1975 the level had fallen to 18.8 mg per cigarette. After 1975 the rate of decrease slowed and in 1988 the average tar content was 13.4 mg per cigarette.



Figure 5.8 Average Tar Yield per Cigarette, UK sales adjusted



### 5.3 Summary

Although tobacco was introduced to Europe in the sixteenth century, cigarette smoking did not become popular until the twentieth century, with women commencing to smoke cigarettes in the 1920s. People tend to start smoking in their late teens and early twenties, with consumption of cigarettes per smoker increasing with age until age 60 years, after which the consumption begins to fall. After the age of 35 years the percentage of the population who smoke begins to fall. This decrease is more marked in the more recent years. Apart for the years of World war II, consumption for males, as measured by cigarettes per male per year, had reached their maximum by 1940. Male consumption remained constant until

the early 1970s, when it began a dramatic decrease. The picture for women is somewhat different in that consumption increased steadily until the early 1970s, whereafter it began a parallel decline to that for males. The data for the percentage of people who smoke is only available from 1948 onwards and again the pattern is different for males and females, The percentage of males who smoke has declined for the whole period with the decline being more rapid after 1970. For females, the percentage who smoke remained fairly constant until 1975, whereafter it followed the male decline. These patterns are reflected in the number of cigarettes smoked per smoker, where the consumption rose for both males and females until 1975 whereafter they began to decline. For the whole time period (1949 to 1987) female smokers have smoked less than their male counterparts.

For both males and females a higher percentage of the younger age group smoke than the older age group. While the percentage of males who smoke in both age groups have fallen over the whole period, the percentage of females in both age groups who smoke did not begin to decline until after 1975. The percentage of females who smoked in the older age group increased between 1955 and 1960 and was not stable as in the younger age group. The patterns of consumption also vary between the sexes. Males smoke more cigarettes per smoker than females, and the older males smoke more than younger males. In both sexes

consumption rose between 1950 and 1975, by approximately 45% in males for both age groups and 150% in females for both age groups. The consumption for males began to decline in 1980, whereas the decline for females is only just beginning.

The composition of cigarettes has also changed over the period of interest. The percentage of smokers who smoke plain cigarettes has dropped from almost 100% in the 1950s to less than 5% in the late 1980s. The tar content of the average cigarette smoked has also declined and this decline has accelerated after 1970, and is now less than 50% of the level in the 1940s.



Chapter 6

**MODELS BASED ON AGE-STANDARDISED**

**LUNG CANCER INCIDENCE AND**

**INFORMATION ON SMOKING BEHAVIOUR**

The previous chapters describe the trends in lung cancer incidence over time in Mersey Region, and the trends in smoking behaviour for Great Britain and the United Kingdom. The projections of lung cancer incidence that are presented in Chapter 3 assume that past trends in lung cancer incidence will remain constant. This implies that there would be no changes in the trends in the risk behaviours associated with lung cancer. The main risk behaviour for lung cancer is known to be cigarette smoking (Doll and Hill 1950). It has been shown that the number of cigarettes smoked per person, the percentage of the population who smoke and the number of cigarettes smoked per smoker began to decrease around the mid 1970s. The consumption of cigarettes had either remained stable (male) or been increasing (female) prior to that time. The composition of cigarettes is changing over time, with a reduction of the average sales adjusted tar content of cigarettes.

This chapter investigates the effect of inclusion of information on smoking behaviour in the projection models. The inclusion of this information should improve the accuracy of the projection models and allow investigation of the effect of changes in smoking behaviour on lung cancer incidence.

This chapter will

---

- determine which smoking variables are most predictive of the age-standardised lung cancer incidence, for each sex separately
  - determine the appropriate lag period between changes in the smoking variables and age-standardised lung cancer incidence, for each sex separately
  - determine the best models for projecting age-standardised lung cancer incidence, for each sex separately
  - estimate future values of the appropriate smoking variables
  - using the best models and appropriate values of the smoking variables, calculate the projected age-standardised lung cancer incidence for 1990, 2000 and 2010
  - assess whether the Health of the Nation targets are achievable
  - determine the future service needs for Mersey Region for lung cancer in 2000
-



## 6.1 Methods for Fitting Models incorporating Information on Smoking Behaviour

The relationships between trends in smoking behaviour and the trends in age-standardised lung cancer incidence rates are assessed in this section. Inherent in these methods is the assumption that the smoking variables that are included in the analysis affect all people in the same way regardless of their age.

The age-standardised rates for lung cancer incidence are calculated for both males and females separately (Section 4.4.1). The standard population is taken as the Mersey Region population for the year 2000 (Section 4.10.1). Only the years 1957 to 1988 are used in the analysis, because of the known improvements in registration rate during the early 1950s.

The smoking variables that are included in the analysis are discussed in Chapter 5 (Table 6.1). All the data used, except the average tar content of cigarettes, are available for each sex separately. The data for the percentage of the population who smoke and the number of cigarettes smoked per smoker are available from 1949 onwards, that for tar content from 1934, and the number of cigarettes per person and the percentage of smokers who smoke plain cigarettes (by interpolation) from 1905 (Table 6.1).

**Table 6.1 Variables Used in the Regression**

<b>Variable</b>	<b>Years Available</b>
Cancer Incidence	1957-1988
Number Cigarettes per Person	1905-1988
Percent Smoking	1949-1988
Number Cigarettes per Smoker	1949-1988
Percentage of Smokers who Smoke Plain Cigarettes	1905-1988 (by interpolation)
Average Tar Content (sales adjusted)	1934-1988

In order to determine the role of the smoking variables in the trends in lung cancer incidence, the log of the age-standardised lung cancer incidence rate is fitted to the log of the smoking variables by unweighted least squares regression (Draper and Smith 1981). Four different lag times between trends in lung cancer incidence and trends in the smoking variables are investigated, i.e. 10 years, 15 years, 20 years and 25 years. Intermediate lag periods are not investigated because of lack of adequate data to investigate single year differences between lag periods.

Data for each lag time are extracted from the basic data sets (Table 6.2). For example, in the 15 year lag investigation smoking data from 1949 to 1973 are used, together with cancer incidence data from 1964 to 1988.

**Table 6.2 Data Used in the Regression Analysis**

<b>Lag Time</b>	<b>Smoking Parameters for years</b>	<b>Cancer Incidence for years</b>
25 yrs	1949 - 1963	1974 - 1988
20 yrs	1949 - 1968	1969 - 1988
15 yrs	1949 - 1973	1964 - 1988
10 yrs	1949 - 1978	1959 - 1988

The models that incorporate only the number of cigarettes per person, average tar content or the percentage of smokers who smoke plain cigarettes smoked are evaluated further. These models are fitted using the longer series of data; i.e. from 1934 for the 25 year lag, from 1937 for the 20 year lag and 1942 for the 15 year lag for both lung cancer incidence and smoking variables. Lung cancer incidence data prior to 1957 are excluded because of the marked improvement in registration following the initiative on the part of the Department of Health in 1954 (Section 3.2). In all these scenarios, except the 25 year lag, cancer data from 1957 to 1988 are used. For the 25 year lag cancer data from 1959 onwards are used. For the 10 year lag the main analysis included data from 1959 and therefore no extended data set is fitted.

The 'best' model for each lag time is determined by forward stepwise regression. That is the smoking variables are initially fitted individually. The



percentage of variation explained by each variable ( $R^2$ ), and the residual mean square error (MSE) are determined for each model to assess the improvement in fit over the null model. For those variables where the decrease in the residual mean square is significant, the other smoking variables are included one by one to determine whether there is any improvement in fit. This is determined by the relative change in the residual mean square (Draper and Smith 1981).

The fit of the models are also assessed by examining plots of the residuals, where the residual is given by:

$$\text{Residual} = \text{Fitted Rate} - \text{Observed Rate}$$

If the fit is adequate the residuals should be randomly distributed when plotted against either the explanatory variable, or the fitted values (Draper and Smith 1981).

If the coefficient for a variable is in an implausible direction, e.g. an increase in smoking leading to a decrease in lung cancer incidence, it is assumed that the relationship, although it may be statistically significant, is nonsensical and it would not be appropriate to base projections upon them.

The 'best' model for each lag time is the model for which each smoking variable contributes significantly to the

fit, the coefficients for the smoking variable are plausible, the residuals are distributed randomly and the residual mean square is the smallest.

In order to compare the projection models for the different lag times, two statistics are evaluated; the percentage of variation explained by the projection equation and the size of the residual mean square.

## 6.2 Results of Fitting the Models

In the analysis of the smoking data from 1949 to 1988 the average tar content of cigarettes contributed significantly to the model for the age-standardised male lung cancer incidence rate for all lag periods except that of 10 years (Table 6.3). The percentage of males who smoke contributed significantly to the model for the 20 year lag. The quantity smoked, either as cigarettes per male or cigarettes per male smoker improved the fit of the model including the average tar content of cigarettes for a 15 year lag. The number of cigarettes smoked per male was the only variable that contributed to the fit of the model for the 10 year lag. For those models incorporating the average tar content of cigarettes and/or the number of cigarettes smoked per male, the findings are not confirmed when the longer data set is used. The residuals for the model for the 25 year lag are not randomly distributed. This would indicate a poor fit.

**Table 6.3 Results of Fitting Male Age Standardised Rates to Smoking Variables**

<b>Variable</b>	<b>Number of Observations</b>	<b>R<sup>2</sup></b>	<b>Residual Mean Square</b>	<b>Intercept (s.e.)</b>	<b>Coefficient (s.e.)</b>
<b>25 Year Lag</b>					
Average Tar (1949-1963)	15	0.50	0.0024	-0.001 (1.404)	1.484 (0.415)
Average Tar (1934-1963)	30	0.02	0.0080	5.785 (1.053)	-0.225 (0.307)
<b>20 Year Lag</b>					
Average Tar (1949-1968)	20	0.76	0.0012	1.776 (0.430)	0.974 (0.128)
Average Tar (1937-1968)	32	0.03	0.0091	5.746 (0.777)	-0.219 (0.226)
% Smokers (1949-1968)	20	0.56	0.0022	1.079 (0.831)	0.976 (0.205)
<b>15 Year Lag</b>					
Average Tar (1949-1973)	25	0.63	0.0015	3.686 (0.218)	0.410 (0.066)
Average Tar (1942-1973)	32	0.00	0.0094	5.044 (0.469)	-0.012 (1.141)



**Table 6.3 (continued) Results of Fitting Male Age Standardised Rates to Smoking Variables**

<b>Variable</b>	<b>Number of Observations</b>	<b>R<sup>2</sup></b>	<b>Residual Mean Square</b>	<b>Intercept (s.e.)</b>	<b>Coefficient (s.e.)</b>
<b>15 Year Lag</b>					
Average Tar	25	0.69	0.0013	0.532 (1.489)	0.460 (0.066)
No. Cigs/Male (1949-1973)					0.363 (0.170)
<b>Average Tar</b>	<b>32</b>	<b>0.20</b>	<b>0.0078</b>	<b>11.020 (2.277)</b>	<b>-0.010 (0.128)</b>
No. Cigs/Male (1942-1973)					-0.725 (0.272)
<b>Average Tar</b>	<b>25</b>	<b>0.76</b>	<b>0.0010</b>	<b>1.318 (0.696)</b>	<b>0.611 (0.079)</b>
No. Cigs/Male Smoker (1949-1973)					0.351 (0.100)
<b>10 Year Lag</b>					
No. Cigs/Male (1949-1978)	30	0.66	0.0032	-5.511 (1.586)	1.280 (0.193)

**Table 6.4 Results of Fitting Female Age Standardised Rates to Smoking Variables**

<b>Variables</b>	<b>Number of Observations</b>	<b>R<sup>2</sup></b>	<b>Residual Mean Square</b>	<b>Intercept (s.e.)</b>	<b>Coefficient (s.e.)</b>
<b>25 Year Lag</b>					
No. Cigs/Female (1949-1963)	15	0.94	0.0017	-2.292 (0.421)	0.849 (0.058)
No. Cigs/Female (1934-1963)	30	0.87	0.0209	-0.517 (0.285)	0.529 (0.041)
<b>20 Year Lag</b>					
No. Cigs/Female (1949-1968)	20	0.95	0.0028	-3.196 (0.395)	0.950 (0.054)
No. Cigs/Female (1937-1968)	32	0.84	0.0253	-2.397 (0.468)	0.833 (0.066)
<b>15 Year Lag</b>					
No. Cigs/Female (1949-1973)	25	0.96	0.0032	-4.422 (0.328)	1.090 (0.044)
No. Cigs/Female (1942-1973)	32	0.92	0.0124	-5.637 (0.485)	1.250 (0.066)
<b>10 Year Lag</b>					
No. Cigs/Female (1949-1978)	30	0.96	0.0051	-5.650 (0.338)	1.229 (0.045)

The residuals for the other models are all randomly distributed. The smallest residual mean squares are obtained for the model incorporating a 20 year lag for the average tar content, and the model incorporating a 15 year lag for the average tar content of cigarettes and the number of cigarettes smoked per male smoker.

The number of cigarettes smoked per female contributed to the model for female lung cancer incidence for all lag periods (Table 6.4). For both the 20 year lag and the 15 year lag there is close agreement of the coefficients for both sets of smoking data. That is the data set constricted to 1949 to 1987 and that including all years. The residuals for the model for the 25 year lag have a U shaped distribution, while for the other lag periods their distribution appears random. As with the males the smaller residual mean squares are obtained with the 20 year lag and the 15 year lag.

The analyses presented above indicate that the best models for fitting male lung cancer age-standardised incidence rates to smoking variables are:

- |                 |   |
|-----------------|---|
| 1: 20 year lag: | Average Tar content of Cigarettes   |
| 2: 15 year lag: | Average Tar content of Cigarettes<br>+ Number of Cigarettes Smoked per<br>Male Smoker |

These two models both explain 76% of the variation in age-standardised lung cancer incidence (Table 6.3). The



coefficient for the 20 year lag indicates that for each percentage decrease in the average tar content there would be an equivalent percent decrease in the male lung cancer incidence. For the 15 year lag model, the coefficients indicate that effect of decreases in the average tar content is not as marked, but that decreases in the number of cigarettes smoked per male smoker also has an effect.

For females, the picture is not quite as clear, because of the smooth exponential increase in the age-standardised rates for female lung cancer. In all the models the number of cigarettes smoked per female is the best predictive smoking variable, and a 15 year lag period gives a marginally better fit than the other lag periods (Table 6.4). The projections that are presented will be based on:

1: 20 year lag	Number of Cigarettes Smoked per Female
2: 15 year lag	Number of Cigarettes Smoked per Female

For all the lag periods the coefficient for the number of cigarettes smoked per female is approximately one. This would imply that for each percentage increase in consumption there is an equivalent percentage increase in lung cancer incidence.

## 6.3 Projections

### 6.3.1 Methods

Section 6.1 describes the investigation of models relating age-standardised lung cancer incidence rates to smoking information. Two questions are addressed: which smoking variables are most predictive and which lag period is the most appropriate. For males, two models are superior; 20 year lag including the average tar content of cigarettes, and 15 year lag including the average tar content of cigarettes and the number of cigarettes smoked per male smoker. These two models are used to project future lung cancer incidence in males. For the females, the number of cigarettes smoked per females fitted best for all the lag periods. The projections based on two lag periods (15 years and 20 years) are presented.

For each model presented the projected age-standardised lung cancer incidence rate, and its confidence interval, for the year 2000 is calculated. The projected rate is determined by using the value of the smoking variable lagged by the appropriate number of years in the projection equation.

For example, the projected age-standardised rate for males based on the average tar content and the number of cigarettes smoked per male smoker, lagged by 15 years is given by:

$$\ln(R^s_{2000}) = 1.318 + 0.611 \ln(\text{tar}_{1985}) + 0.351 \ln(\text{CS}_{1985})$$

where  $\text{tar}_{1985}$  is the average tar content of cigarettes in 1985

and  $\text{CS}_{1985}$  is the number of cigarettes smoked per male smoker in 1985

The confidence interval is given by:

$$R^s_{2000} \pm t_{\nu, 0.05} s(R^s_{2000})$$

where  $s(R^s_{2000}) = \text{sqrt}(\{\text{MSE}\} \times \{1 + X_0(X'X)^{-1}X_0'\})$

$\nu$  is the degrees of freedom

and  $X$  is the design matrix for the regression

$X_0$  is the matrix of explanatory variables for the year of projection

Projected rates are calculated for the years 1990, 2000, and 2010, using the four projection equation described above. Two lag periods are used; 15 years and 20 years. Therefore smoking data for 1995 are required for the projections for the year 2010. This data is not yet available, and is estimated through linear projection, against year, of the trends in the smoking variables. Data from the years 1976 to 1987 are used for this analysis, because over this period the trends in the



smoking variables are fairly linear (Figures 5.1, 5.5 and 5.8).

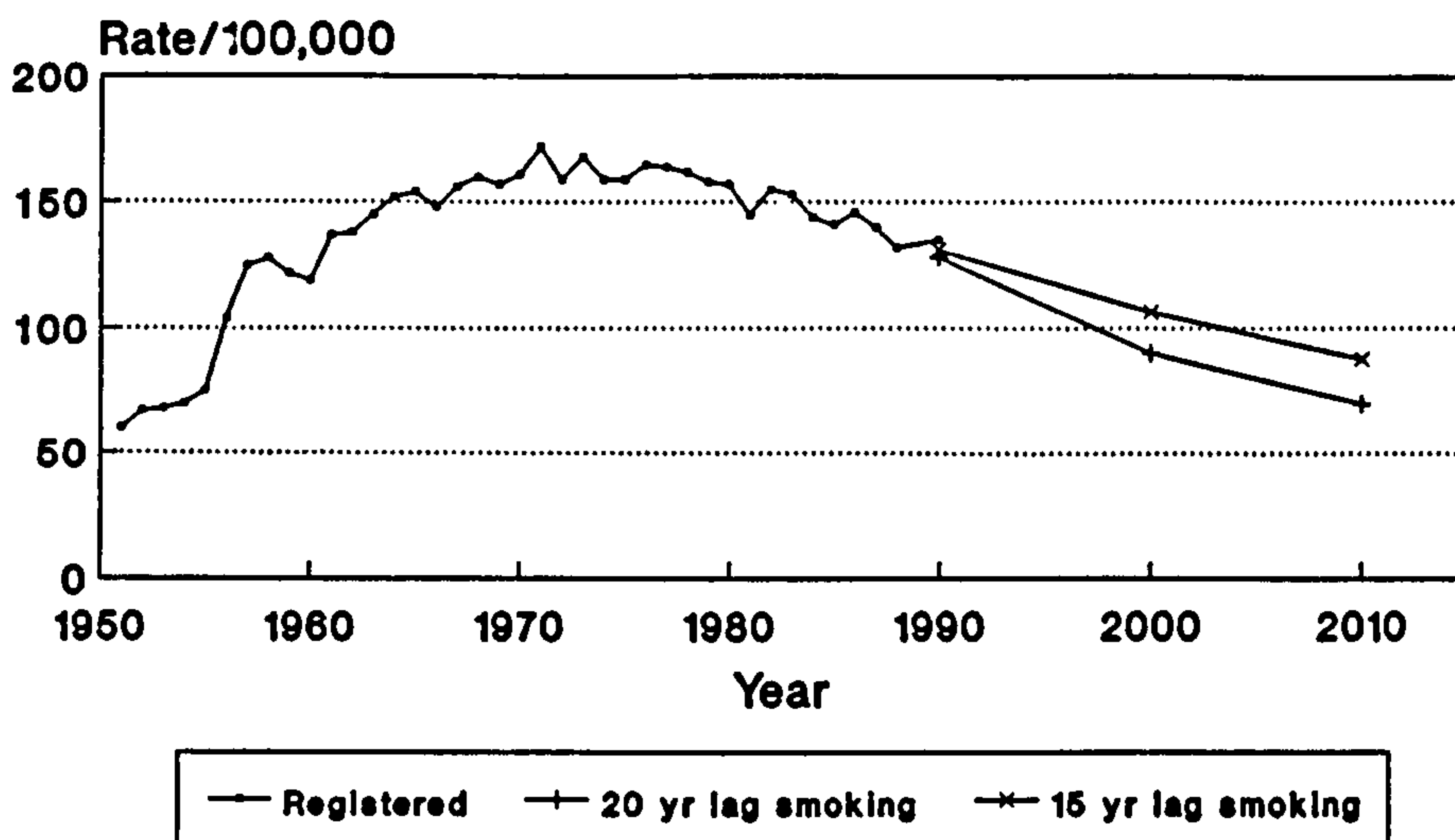
The projected age-standardised incidence rates for 1990 are compared to the age-standardised rates observed in 1990, to assess the accuracy of the projections. The projections are also compared to those obtained by linear extrapolation of the lung cancer incidence trends. The projected rates for 2010 are used to evaluate the potential to achieve the targets set by the Health of the Nation. Service needs are estimated for the year 2000 using the projected number of incident lung cancer cases in that year and data on service needs from 1983-87.

### **6.3.2 Results**

Using the models discussed above (Section 6.2) the age-standardised incidence rates for lung cancer are projected to decline for both males and females (Figure 6.1 and 6.2, Appendices VI(1) and VI(2)). This decline reflects the decreasing consumption of cigarettes in recent years for both males and females and the decreasing average tar content of cigarettes.

The projected age-standardised incidence rates for 1990 are similar to those observed, for both males and females. This would indicate the projection models are adequate. The decline in the rates for males are steeper

**Figure 6.1 Projected Age-Standardised Male Lung Cancer Incidence Rates, using Models incorporating Smoking Information**

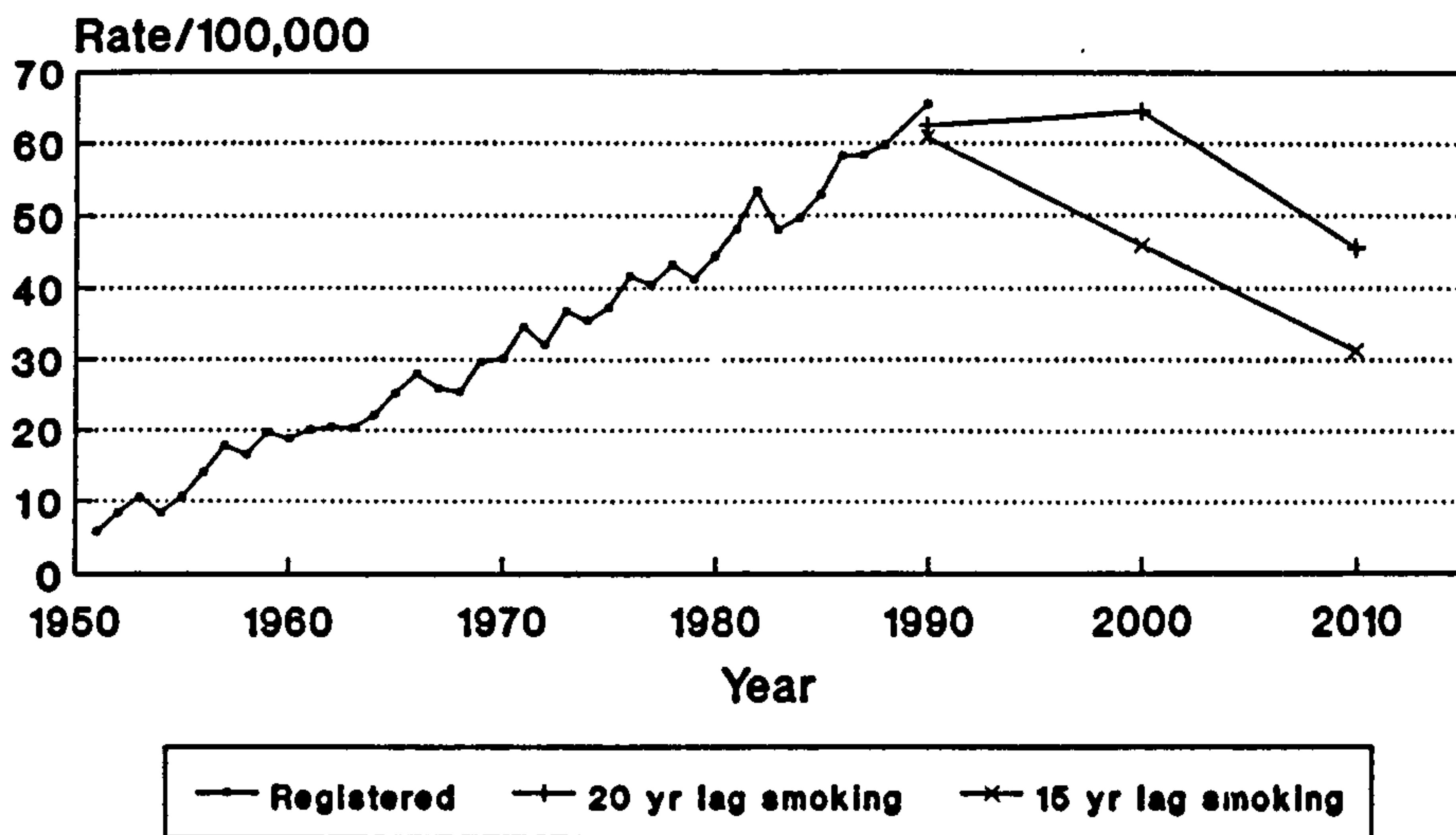


Standardised to 2000 MRHA Population  
Mersey Regional Cancer Registry

than those projected on year alone (Figure 4.10). This reflects the accelerated decrease in the average tar content of cigarettes after 1970, and also the change in the trend of cigarettes smoked per male smoker around 1975 (Figures 5.5 and 5.8). The change in the trend in average tar content has a more marked effect than the number of cigarettes smoked per male smoker on the projections. The projected age-standardised incidence rates in 2010 are 70.0 cases per 100,000 for the model including average tar content with a 20 year lag period, and 87.0 cases per 100,000 for the model including average tar content and the number of cigarettes smoked per male smoker with a 15 year lag.

The results for the females are somewhat surprising, in that a decline in age-standardised lung cancer incidence

**Figure 6.2 Projected Age-Standardised Female Lung Cancer Incidence Rates, Models incorporating Smoking Information**



Standardised to 2000 MRHA Population  
Mersey Regional Cancer Registry

is projected to occur over the next twenty years (Figure 6.2). This is contrary to the present trend in female lung cancer where a 100% increase is projected to occur (Figure 4.11). The projected decrease in cancer incidence is a reflection on the decrease in the number of cigarettes smoked per female that occurred after 1975. The effect of this change smoking behaviour depends upon the lag time used in the model. As explained in Section 6.2, the consistent exponential increase in female age-standardised lung cancer incidence make it difficult to differentiate among lag times.

The changes in the trends in smoking variables that have been included in these models are assumed to affect all age groups in the same way at the same time. However, cigarette consumption, as measured either by cigarettes



smoked per person or per smoker, is usually determined by the birth cohort. That is people who are born in the same era would tend to have similar smoking patterns.

Therefore, population decreases or increases in cigarette consumption would affect different birth cohorts of the population differently, with the changes affecting the younger cohorts more. The older cohort are likely to be less affected by these changes. Since lung cancer is a disease of the elderly, the projected age-standardised lung cancer incidence is likely to be overestimated.

### 6.3.3 Health of the Nation Targets

Age-standardised lung cancer incidence in males is projected to decrease from 134.6 cases per 100,000 in 1990 to 70.0 cases per 100,000 for the model with a 20 year lag, or 87.0 cases per 100,000 for the model with a 15 year lag. These projection would represent a 48% and a 35% decrease for the two models respectively. Thus if these projection are appropriate the Health of the Nation targets will be met for males, assuming present trends in the average tar content of cigarettes, and the number of cigarettes smoked per male smoker continue.

The Health of the Nation target for the reduction of female lung cancer by 15% will also be met. The projected age-standardised female lung cancer incidence rates are 45.6 cases per 100,000 for the model with a 20

year lag and 31.3 for the model with a 15 year lag. These represent reductions in female lung cancer incidence between 1990 and 2010 of 30% and 52% respectively. If the lag period were 25 years this decrease would not have commenced. As with the projections based on the extrapolation of trends in the age-standardised lung cancer incidence over calendar period, these comparisons are based on lung cancer incidence in those under 85 years of age (Section 4.10.3).

#### 6.3.4 Service Needs

Sex	Lag Period	Surgery	Chemo-therapy	Radio-therapy
Male	20 yrs	59	58	109
Female	20 yrs	84	82	155
Total	20 yrs	125	140	264
Male	15 yrs	41	40	75
Female	15 yrs	60	58	110
Total	15 yrs	101	98	185

Incorporating information on smoking behaviour gives projected service needs that are substantially lower than those based on the extrapolation of trends in lung cancer incidence (Table 6.5). These projections indicate that

service needs could be up to 50% lower than in the mid 1980s. However as explained earlier (Section 4.10.4) these projections must be interpreted with caution because they are not adjusted for the age distribution of the incident cancer cases.

#### 6.4 Summary

This chapter has investigated the relationships between age-standardised lung cancer incidence and the smoking variables. The best models obtained are:

<b>Male</b>	
1: 20 year lag:	Average Tar content of Cigarettes
2: 15 year lag:	Average Tar content of Cigarettes + Number of Cigarettes Smoked per Male Smoker
<b>Female</b>	
1: 20 year lag	Number of Cigarettes Smoked per Female
2: 15 year lag	Number of Cigarettes Smoked per Female

Projections based on these models indicate that lung cancer incidence will decline for both males and females. The projected decline in males is steeper than that projected on year alone. The result for females is surprising, but is related to the decline in the number of cigarettes smoked per female that started in the mid



1970s. The length of the lag period determines the extent of the decline.

The decrease in lung cancer incidence is adequate to meet the Health of the Nation targets. However, the models assume that the changes in smoking behaviour affect all ages in the same fashion. It is generally believed that smoking habits are acquired early in life (Hammond 1966) and are determined by birth cohort (Doll and Hill 1964). Therefore the effect on lung cancer incidence of decreases in the quantity smoked is probably less than that projected in these models. The decreases in quantity affecting the younger cohorts more than the older cohorts.

The service needs are also projected to be lower than those projected on year alone. These projections are based on the total number of lung cancer cases. As explained above, the projected total number of lung cancer cases may be too low. Also, the estimates of service needs are not adjusted for the differences in treatment according to age.

Chapter 7

**MODELS BASED ON BROAD-BAND**

**AGE-SPECIFIC LUNG CANCER**

**INCIDENCE RATES AND INFORMATION**

**ON SMOKING BEHAVIOUR**

The last chapter investigates the relationship between the trends in age-standardised lung cancer incidence rates and trends in variables relating to smoking behaviour. Four different lag times between smoking behaviour and the development of lung cancer are also investigated. Projections are based on the 'best' models.

This chapter will address the same questions, but in relation to the broad-band age-specific lung cancer incidence rates. Two broad-band age groups will be used; 45-64 years and 65+ years of age. The younger age group, 30-44 years, is not included in these analyses, because of the small number of cancer cases occurring in this age group and the relative stability of the trend. For the middle age group (45-64 years), the data on quantity smoked and the percentage of the population who smoke are given for the age group 25-34 years, while for the older age group (65+ years) the equivalent data is for the 35-59 year age group (Section 7.1). Lag periods of 10 years, 15 years, 20 years and 25 years are investigated.



For each sex and broad-band age group this chapter will:

---

- determine which smoking variables are most predictive of the broad-band age-specific lung cancer incidence
  - determine the appropriate lag period between changes in the smoking variables and lung cancer incidence
  - determine the best models for projecting lung cancer incidence
  - estimate future values of the appropriate smoking variables
  - calculate the projected broad-band age-specific lung cancer incidence, using the best models and appropriate values of the smoking variables
  - compare the projected broad-band age-specific rates for 1990 with those observed
  - assess whether the Health of the Nation targets are achievable
  - determine the future service needs for Mersey Region for lung cancer in 2000
-

## 7.1 Methods for Fitting Models incorporating Information on Smoking Behaviour

The number of lung cancer cases that occur each year in the broad-band age groups can be assumed to be Poisson distributed (Section 4.11.1).

The model relating lung cancer incidence to information on smoking behaviour is given by:

$$\ln(R) = \ln(\mu) - \ln(N) = X\beta$$

where  $R$  is the vector of age-specific incidence rates  $R_i$

$\mu$  is the vector of age-specific frequencies  $\mu_i$

$N$  is the vector of age specific populations  $N_i$

$X$  is the design matrix, and

$\beta$  is the vector of coefficients

The design matrix includes data on the smoking variables. The hypothesis is that the broad-band age-specific rates are dependant on the smoking variables lagged by an appropriate period. Lag times of between 10 and 30 years give the highest correlation between smoking behaviour and lung cancer incidence (Hakama and Pukkala, 1984). In the 45-64 year age group the median age for cancer incidence is 58 years and in the 65+ year age group the median age for cancer incidence is 72 years Thus in these analyses, the smoking habits of the age group 25-34 years

should be correlated with the lung cancer incidence in the 45-64 age group, and smoking habits in the age group 35-59 years should be correlated with incidence in the 65+ age group.

The annual broad-band age-specific lung cancer incidence rates for male lung cancer in the 65+ years age group are fitted against the number of cigarettes smoked per male per week, the percentage of the male population who smoke and the number of cigarettes smoked per male smoker per week, in the 35-59 year age group; the percentage of male smokers who smoke plain cigarettes and the average tar content of cigarettes. A similar analysis is carried out for females in the same age group using the data for females where available.

In the younger age group, 45-64 years, the broad-band age-specific incidence rates for males are fitted against the number of cigarettes smoked per male per week, the percentage of the male population who smoke, the number of cigarettes smoked per male smoker per week, in the 25-34 year age group, the percentage of male smokers who smoke plain cigarettes and the average tar content of cigarettes. A similar analysis is carried out for females in the same age group using the data for females where available.

The data are modelled using GLIM software (Baker and



Nelder 1978), which allows the fitting of Poisson distributed data (Section 4.11.1). The log of the number of incident cancer cases is fitted to the logs of the smoking variables using forward stepwise regression. The contribution of each variable to the model is determined by examination of the change in the scaled deviance ( $S_{M_2;M_1}$ ) (Section 8.1, McCullagh and Nelder 1989), which is given by:

$$S_{M_2;M_1} = -2l_2 + 2l_1$$

where	$M_1$	denotes Model 1 based on $n_1$ parameters
	$M_2$	denotes Model 2 based on $n_2$ parameters and the parameters of model 1 are included in model 2
	$l_1$	is the log likelihood for model 1
	$l_2$	is the log likelihood for model 2

The distribution of  $S_{M_2;M_1}$  is approximated by that of a chi square with  $(n_2 - n_1)$  degrees of freedom. Thus the contribution of each variable can be evaluated by comparing the change in the scaled deviance with the chi squared distribution with the appropriate degrees of freedom.

Goodness of fit of the model can be determined by the final scaled deviance, which again is approximately distributed as a chi square with appropriate degrees of

freedom. The distribution of the residuals also should be examined. For these analyses the appropriate residual would be the Pearson residual (McCullagh and Nelder 1981), which is defined as:

$$r_x = (I_x - \bar{I}_x) / \sqrt{\bar{I}_x}$$

where  $I_x$  is the observed count for data point  $x$   
and  $\bar{I}_x$  is the predicted count for data point  $x$

If the model is appropriate then the residuals should show a random distribution when plotted against the fitted value. However, examination of the plots may reveal a systematic pattern in the residuals, an overall wide dispersion, or one or more outliers. If the residuals show a systematic pattern when plotted against the covariates, the model may be inappropriate. Inclusion of other smoking variables may improve the fit of the model.

If residuals show no systematic pattern but the final scaled deviance is large then it may be assumed that the data is over dispersed. This could be due to the cancer incidence in Mersey Region, being distributed as a clustered Poisson processes (Breslow 1984). If the data are over dispersed the estimates of the regression coefficients are unbiased, and the standard errors for these estimates are obtained by multiplying those obtained from the Poisson model by an estimate of the dispersion parameter. This is a simple scale factor

calculated from the square root of the final scaled deviance divided by the degrees of freedom (McCullagh and Nelder 1989).

Outliers must be examined individually to determine if there is a problem with the original data.

Four lag periods are investigated; 25 years, 20 years, 15 years and 10 years. The ranges of years for the data included in these analyses are the same as for the age-standardised analyses (Table 6.2).

In this manner the 'best' models for each sex and age group are determined. Where the same variables are included in the 'best' models for different sexes, ages or lag periods, the parameter estimates are compared to investigate consistency of the models.

## **7.2 Results of Fitting the Models**

The average tar content per cigarette contributed significantly to the model for males aged 65 years and over with lag periods of 25 years, 20 years and 10 years (Table 7.1). The number of cigarettes is the only variable to contribute significantly to the model for the 15 year lag period. The number of cigarettes smoked per male smoker improved the fit of the model incorporating the average tar content of cigarettes for the 10 year lag



period.

Table 7.1 Results of Fitting Broad-Band Age-Specific Lung Cancer Incidence Rates to Smoking Variables for Males aged 65+ Years

Parameter	Degrees of Freedom	Scaled Deviance	Intercept (s.e.)	Coefficient (s.e.)
<b>25 year lag</b>				
Average Tar	13	22.15	-2.685 (0.904)	0.708 (0.267)
<b>20 year lag</b>				
Average Tar	18	32.85	-1.221 (0.403)	0.273 (0.120)
<b>15 year lag</b>				
No. Cigs/Male	23	57.09	-1.703 (0.540)	0.318 (0.122)
<b>10 year lag</b>				
No. Cigs/Male Smoker +	27	151.21	-9.252 (0.897)	1.388 (0.124)
Average Tar				0.364 (0.058)

The final scaled deviances are significant for all lag periods except the 25 year lag. The residuals for the 25 year lag have a  $\cap$  shaped distribution. Those for the 10 year lag show a positive correlation with the fitted values. The residuals are distributed randomly for the 20 year lag and the 15 year lag. This indicates that for these lag periods the data are over dispersed and that the fit is adequate. Thus the best models for the males age 65 years and over are that incorporating a 20 year lag for the average tar content of cigarettes and that incorporating a 15 year lag for the number of cigarettes

smoked per male (Table 7.5).

Table 7.2 Results of Fitting Broad-Band Age-Specific Lung Cancer Incidence Rates to Smoking Variables for Males aged 45-64 Years

Parameter	Degrees of Freedom	Scaled Deviance	Intercept (s.e.)	Coefficient (s.e.)
<b>25 year lag</b>				
% who Smoke	13	16.11	-9.647 (0.873)	1.929 (0.208)
<b>20 year lag</b>				
% who Smoke	17	17.76	-6.570 (0.746)	1.076 (0.224)
% Male Smokers who Smoke Plain Cigs				0.134 (0.053)
<b>15 year lag</b>				
Average Tar	22	28.67	-3.252 (0.060)	0.347 (0.239)
% Male Smokers who Smoke Plain Cigs				0.147 (0.049)
<b>10 year lag</b>				
Average Tar	27	49.73	-7.584 (0.866)	0.932 (0.084)
No. Cigs/Male Smoker				0.626 (0.127)

The percentage of males who smoke contributed significantly to the models for males aged 45-64 years for the lag periods 25 years and 20 years (Table 7.2). The addition of the percentage of male smokers who smoke plain cigarettes improved the fit for the 20 year lag. The average tar content of cigarettes contributed

significantly to the models for the lag periods 15 years and 10 years. For the 15 year lag the addition of the percentage of male smokers who smoke plain cigarettes improved the fit. For the 10 year lag the addition of the number of cigarettes smoked per male smoker improved the fit.

The final scaled deviances are not significant except for the model for the 10 year lag. For this lag period the residuals show increased variation with increasing fitted values. The residuals for the models for both the 20 year lag and the 15 year lag show a similar pattern. Therefore the best model for the males aged 45-64 years is based on a 25 year lag for the percentage of males who smoke (Table 7.5).

For the females aged 65 years and older, the number of cigarettes smoked per female contribute significantly to the models for all lag periods, except for the 10 year lag (Table 7.3). In these models the parameter estimates for the number of cigarettes smoked per female are approximately equal. For the ten year lag period the number of cigarettes smoked per female smoker contributes significantly to the model. The addition of the percentage of female smokers that smoke plain cigarettes improves the fit for models for the 20 year, 15 year and 10 year lag periods. The parameter estimates for this variable are consistent for these three lag



periods. However, the parameter estimates are all negative, where they would be expected to be positive.

Table 7.3 Results of Fitting Broad-Band Age-Specific Lung Cancer Incidence Rates to Smoking Variables for Females aged 65+ Years

Parameter	Degrees of Freedom	Scaled Deviance	Intercept (s.e.)	Coefficient (s.e.)
<b>25 year lag</b>				
No. Cigs/ Female	12	21.73	-4.895 (0.218)	0.870 (0.061)
<b>20 year lag</b>				
No. Cigs/ Female	17	31.10	-3.599 (0.402)	0.640 (0.086)
+ % Female Smokers who Smoke Plain Cigs				-0.163 (0.028)
<b>15 year lag</b>				
No. Cigs/ Female	22	27.46	-3.845 (0.416)	0.643 (0.935)
+ % Female Smokers who Smoke Plain Cigs				-0.160 (0.023)
<b>10 year lag</b>				
No. Cigs/ Female Smoker	27	41.39	-6.985 (0.777)	1.163 (0.154)
+ % Female Smokers who Smoke Plain Cigs				-0.132 (0.028)

The distribution of the residuals for all the lag periods except 20 years appear to be random. For the 20 year lag period the residuals show a non random pattern, possibly due to a high registration rate in 1982.

Table 7.4 Results of Fitting Broad Band Age Specific Lung Cancer Incidence Rates to Smoking Variables for Females aged 45-64 years

Parameter	Degrees of Freedom	Scaled Deviance	Intercept (s.e.)	Coefficient (s.e.)
<b>25 year lag</b>				
No. Cigs/ Female	12	10.85	-2.948 (1.022)	0.419 (0.195)
+ % Females Smokers who Smoke Plain Cigs				-0.251 (0.094)
<b>20 year lag</b>				
No. Cigs/ Female Smoker	18	15.92	-7.889 (0.473)	1.213 (0.109)
<b>15 year lag</b>				
No. Cigs/ Female Smoker	23	78.08	-8.273 (0.329)	1.255 (0.074)
<b>10 year lag</b>				
No. Cigs/ Female Smoker	27	81.32	-4.667 (0.833)	0.539 (0.161)
% Female Smokers who Smoke Plain Cigs				-0.1814 (0.0350)

The apparent lack of differentiation among the lag periods may be due to the constant increase of the log rates over time. Therefore projections are based on the models for the 25 year lag period, and the 15 year lag period (Table 7.5).

The number of cigarettes per female smoker contributes significantly to the models for females aged 45-64 years for all lag periods except 25 years (Table 7.4). The

number of cigarettes smoked per female contributes significantly to the model for the 25 year lag. The addition of the percentage of female smokers who smoke plain cigarettes improves the fit of the models for the 25 year lag and the 10 year lag. However, the coefficient for this variable is negative for both lag periods, indicating a non-plausible relationship.

The final scaled deviance is significant for the lag periods of 15 years and 10 years. The residuals are distributed randomly for the model for the 25 year lag. For the 20 year lag the residuals show increasing variation with increasing fitted values, and for the 15 year and 10 year lag the residuals have a  $\cap$  shaped distribution.

Thus, there is no model for females aged 45-64 years that fits the data adequately and gives plausible coefficients for all the smoking variables included in the model. The only model for which the residuals are distributed randomly is that including a 25 year lag for the number of cigarettes smoked per female and the percentage of female smokers who smoke plain cigarettes. This is the model which is used for projections (Table 7.5).



**Table 7.5 The 'Best' Models for each Sex and Age Group**

**Male 65+ years**

20 year lag      Average Tar

15 year lag      Number of Cigarettes Smoked per Male

**Male 45-64 years**

25 year lag      Percentage of Males who Smoke

**Female 65+ years**

25 year lag      Number of Cigarettes Smoked per Female

15 year lag      Number of Cigarettes Smoked per Female  
+ Percentage of Female Smokers who Smoke  
Plain Cigarettes

**Female 45-64 years**

25 year lag      Number of Cigarette Smoked per Female  
Smoker  
+ Percentage of Female Smokers who Smoke  
Plain Cigarettes

**7.3 Projections**

**7.3.1 Methods**

Future lung cancer incidence is projected for each of the models identified in the model fitting. The number of lung cancer cases for the years 1990, 2000 and 2010 are determined from the projection equations using the values of the smoking variables for the appropriate years given the different lag periods. Thus for a 15 year lag, smoking data from 1985 would be used for projections in

the year 2000. Similarly for a 20 year lag, smoking data from 1980 would be used for projection to the year 2000.

The variance of the projected incidence is calculated using the method outlined by Hakulinen and Dyba (1994):

$$\text{var}(I_s) = \text{var}(\mu) + E(\mu)$$

GLIM provides the variance of  $\ln(\mu)$  and using the approximation

$$\text{var}(\mu) = \mu^2 \text{var}(\ln(\mu))$$

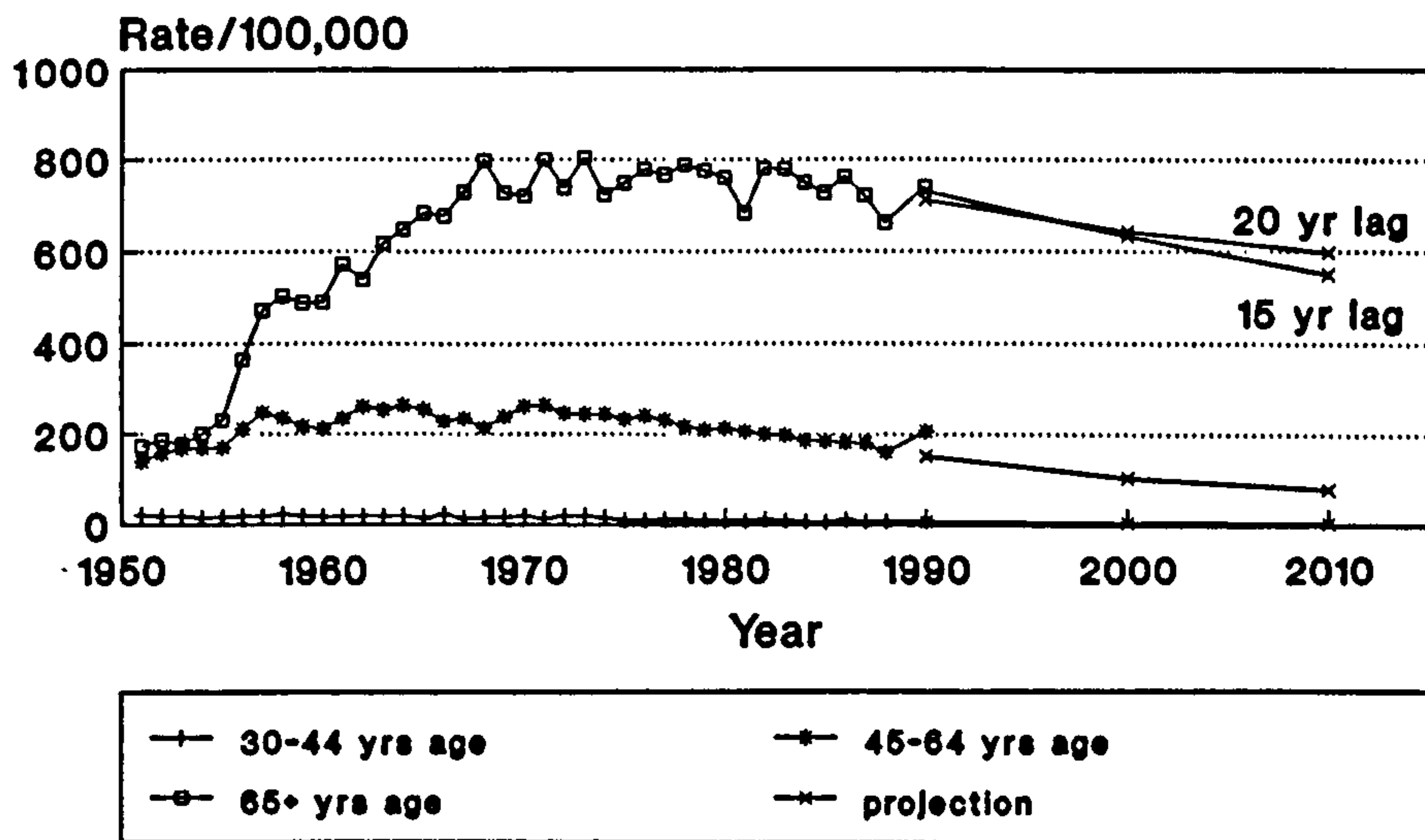
the variance of the projected rate can be calculated, and hence confidence intervals estimated.

As with the age-standardised incidence rates, the projected broad-band incidence rates for 1990 are compared with those observed. The projections for 2010 are used to assess whether the Health of the Nation targets will be met, and the projections for the year 2000 will be used to assess service needs.

### 7.3.2 Results

The age-specific lung cancer incidence rates for males are projected to decrease for both age groups (Figure 7.1). These rates of decrease are larger than those for the projections based on linear extrapolation of the past trends. This is due to the reduction in smoking that has

**Figure 7.1 Projected Male Broad-Band Age Specific Lung Cancer Incidence Rates, Models incorporating Smoking Information**



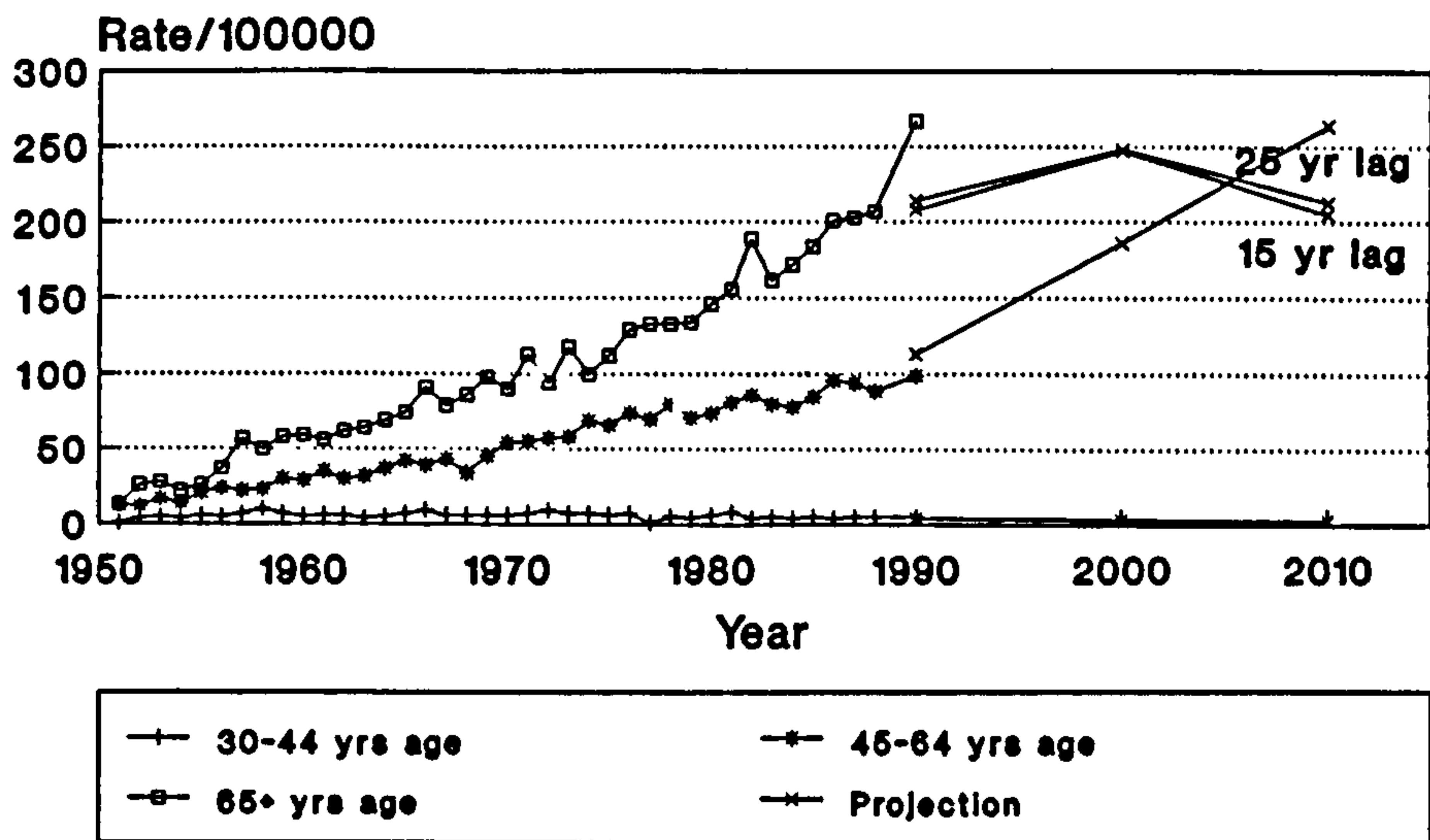
Mersey Regional Cancer Registry

occurred over the last 20 years. The confidence intervals for the projections based on models incorporating information on smoking behaviour are generally wider for the those based on extrapolation of past trends (Appendix VII).

For females the picture appears to be different. The lung cancer incidence in the older age group appears to peak around the turn of the century and then start to decline (Figure 7.2). Again this change in trend is reflecting the decrease in smoking that is occurring among females, particularly in the older age groups. The picture for the middle age females is less optimistic, with markedly increasing lung cancer burden over the next 20 years. This is a reflection of the model used: the trends in



**Figure 7.2 Projected Female Broad-Band Age-Specific Lung Cancer Incidence Rates Models incorporating Smoking Information**



Mersey Regional Cancer Registry

lung cancer incidence are related to the number of cigarettes smoked per female smoker. Although the percentage of females who smoke has decreased in recent years the number of cigarettes smoked per female smoker has increased. The model also includes the percentage of female smokers who smoke plain cigarettes, but lung cancer incidence is inversely related to this variable. Thus any decrease in this percentage would increase the estimated cancer burden. The picture for females, when smoking information is included in the model is very different from those projections based on past trends alone (Figure 4.13), where the lung cancer incidence is projected to increase exponentially in both age groups, but the rate of increase being higher in the older age group. The confidence intervals for the projections based on models incorporating information on smoking behaviour

are wider than those based on extrapolation of past trends for females age 65 years and over, but are approximately the same for the females age 45-64 years (Appendices VII and IV(4b)).

When the projected rates are compared with those actually observed, the agreement is as good as for the projections based on past trends alone (Figure 4.12, 4.13, 7.1 and 7.2).

### 7.3.3 Health of the Nation Targets

These projections indicate that the Health of the Nation targets will only be met for males in the 45-64 year age group, where a 48% reduction in lung cancer incidence is projected. For the older males the projected reduction is 16% for the model incorporating a 20 year lag, and 25% in that incorporating a 15 year lag. For females the picture is worse. For the older females the projected lung cancer incidence rates for 1990 and 2010 are projected to be very similar, for both the models. For the middle age group the lung cancer incidence is projected to more than double between 1990 and 2010.

The attainment of the Health of the Nation targets is based on changes in smoking behaviour occurring before the year 2000. The models that are used for these projections are mostly based on lag periods of 20 years

or 25 years. Thus the smoking behaviour in 1985 or 1990 would determine the lung cancer incidence for the year 2010. The Health of the Nation targets for changes in smoking behaviour by the year 2000 would not affect cancer incidence in 2010.

#### 7.3.4 Service Needs

**Table 7.6 Service Need Projected for 2000 using Broad-Band Age-Specific Incidence Rates and Smoking Information**

<b>Age</b>	<b>Surgery</b>	<b>Chemotherapy</b>	<b>Radiotherapy</b>
<b>Male</b>			
45-64 years	50	44	74
65+ years			
20 year lag	62	58	142
15 year lag	61	57	139
<b>Total</b>	<b>111/112</b>	<b>101/102</b>	<b>213/216</b>
<b>Female</b>			
45-64 years	84	95	132
65+ years			
25 year lag	29	38	78
15 year lag	29	38	78
<b>Total</b>	<b>113</b>	<b>133</b>	<b>210</b>

If the models incorporating smoking behaviour are used to project service needs for the year 2000, approximately 225 cases will require surgery, 235 cases will require



chemotherapy and 425 cases will require radiotherapy (Table 7.6). These total are very similar to the projected service needs using linear extrapolation of the trends over years. The major difference in the projected rates between the two methods of projections occurred in the older age group where only a small percentage receive specific treatment (Appendix III(3)). Also the projections based on the two lag periods for the older age groups for both males and females do not differ substantially. Service needs are dominated by the middle age females.

#### 7.4 Summary

This chapter has examined the possible improvements in projecting broad-band age-specific lung cancer incidence by incorporating information on smoking behaviour in the projection models. For both males and females separately, aged 65 years and older, there is no single model that is superior. The best models for lung cancer incidence among males are that incorporating a 20 year lag for the average tar content of cigarettes, and that incorporating a 15 year lag for the number of cigarettes smoked per male. For females the best models are that incorporating a 25 year lag for the number of cigarettes smoked per female and that incorporating a 15 year lag for the percentage of female smokers who smoke plain cigarettes. The best model for males aged 45-64 years incorporates a

25 year lag for the percentage of males who smoke. For the females in this age group, no plausible model fits the data adequately. The model incorporating a 25 year lag for both the number of cigarettes smoked per female smoker and the percentage of female smokers who smoke plain cigarettes is the model chosen for the projection. In both the models for females incorporating the percentage of female smokers who smoke plain cigarettes, the coefficient of the parameter estimate for this variable is negative.

Projections based on these models for 1990 agree closely with the observed broad-band age-specific rates. The Health of the Nation targets will be achieved only by the males aged 65 years and older. The incidence rates in the females aged 45-64 years are projected to double between 1990 and 2010. The projections for service needs indicate that the females aged 45-64 years will require most of the services. However, the total service needs projected for the year 2000 do not differ markedly from those projected using models based on extrapolation over year.

Chapter 8

**AGE-PERIOD-COHORT MODELLING**



In chapters 4, 6 and 7 age-specific rates for lung cancer in Mersey Region have been examined in two ways. These were summary statistics, such as the age standardised rates and the broad-band age-specific rates, and by visual examination of age-period and age-cohort specific rates. The summary statistics used reflect trends in cancer incidence over calendar period. The changes in the trends of the risk factors for lung cancer incidence over calendar period are assumed to affect all age groups in a similar fashion. However, as discussed earlier smoking habits are acquired early in life and trends in such variables as quantity smoked are thought to vary with birth cohort rather than calendar period (Section 2.2.3).

Whether the trends over time are due to changes over calendar period or birth cohorts can be investigated by visual examination of age-period and age-cohort plots. This chapter will present models which allow the determination of the relative contribution of each of these effects to the variation in rates by quantifying changes in age-specific rates over calendar period and/or birth cohort.

In particular this chapter will :

---

- describe the Age-Period-Cohort model
  - describe the hierarchy for determining which time effects contribute to the trends in incidence rates
  - determine the 'best' models for lung cancer incidence for males and females
  - estimate the age, period and cohort parameters necessary for projections of lung cancer incidence for the quinquennia centred on 1991, 2001 and 2011
  - calculate projected age-specific incidence rates for the quinquennia centred on 1991, 2001 and 2011, using the 'best' models and appropriate parameter estimates
  - compare the projected age-specific lung cancer incidence for the quinquennium centred on 1991 with that observed
  - assess whether the Health of the Nation targets are achievable
  - determine the service needs for Mersey Region for the quinquennia centred on 2001
-

## 8.1 Methods for Age-Period-Cohort Variation

The basic model described in this chapter assumes that the age-specific lung cancer incidence rates are dependant on an age effect and/or a calendar period effect and/or a birth-cohort effect. The full model is:

$$\ln(R_{ijk}) = \ln(I_{ijk}) - \ln(N_{ijk}) = \alpha_i + \beta_j + \gamma_k$$

where  $i$  is the indicator for age group  
 $i=1,2,\dots,I$

$j$  is the indicator for calendar-period  
 $j=1,2,\dots,J$

$k$  is the indicator for birth-cohort  
 $k=1,2,\dots,K$

$R_{ijk}$  is the expected age-specific incidence rate for age group  $i$ , calendar-period  $j$  and birth-cohort  $k$

$I_{ijk}$  is the age-specific incidence frequency for age group  $i$ , calendar-period  $j$  and birth-cohort  $k$

$N_{ijk}$  is the age-specific population frequency for age group  $i$ , calendar-period  $j$  and birth-cohort  $k$

$\alpha_i$  is the age parameter for age group  $i$

$\beta_j$  is the calendar-period parameter for calendar-period  $j$

$\gamma_k$  is the birth-cohort parameter for birth-cohort  $k$



It should be noted that the indicators for age group, calendar-period and birth-cohort are related; viz

$$k = I - i + j$$

The model can be expressed in vector notation:

$$\ln(R) = \ln(\mu) - \ln(N) = X\beta$$

where  $X$  is the design matrix of indicator variable for age group, calendar-period and birth-cohort,  
and  $\beta$  is the vector of coefficients  $\alpha_1$ ,  $\beta_j$  and  $\gamma_k$

The data used for these investigations are the five year age-specific frequencies for male and female lung cancer for the quinquennia centred on the census years (1961, 1971, 1981) and the intervening years when 10% samples of the population are taken (1966, 1976, 1986). Person years at risk are estimated by multiplying the five year age-specific population frequencies for the given years by five. Thus data are available for the six quinquennia from 1961 to 1986, for the five year age groups. Given the small number of lung cancer cases in males less than 30 years and females less than 35 years (less than 20 cases per quinquennia); and the lack of good population data and the low registration rates for those aged 84 year and older (Williams et al 1993), the data used in the analyses are restricted to the age ranges 30 to 84 years for males and 35 to 84 years for females (Tables 4.1 and 4.2).

As described earlier (Section 4.11.1) the number of incident cancer cases occurring in each age group for each quinquennia have a Poisson distribution and thus the log likelihood is of the form (McCullagh and Nelder 1989) :

$$l(\mu, I) = \sum_{ij} (I_{ij} \ln \mu_{ij} - \mu_{ij})$$

where  $\mu_{ij}$  is the expected age-specific frequency for age group  $i$  and calendar-period  $j$

and  $I_{ij}$  is the observed age-specific frequency for age group  $i$  and calendar-period  $j$

The likelihood measures the support provided by the data for the particular values of the parameter estimates of the probability model (Clayton and Hills 1993). It is calculated by determining how probable the observations would be if the parameters were to have the assumed values. The method of maximum likelihood estimation maximises this likelihood, or more commonly the log of the likelihood for the given set of data.

The simplest form of the model assumes that there are no changes in the age-specific rates over calendar period or birth cohort and all variation is due to the age effects. i.e.

		$\ln(R) = \ln(\mu) - \ln(N) = X_A \alpha$
where	$X_A$	is the matrix $\{A_1, A_2, \dots, A_i, \dots, A_I\}$
	$\alpha$	is the vector of age parameter estimates
	$A_i^T$	$= \{0_J, 0_J, \dots, 1_J, \dots, 0_J\}$
	$0_J$	is a vector of length J of zeros
	$1_J$	is a vector of length J of ones.

The age parameters  $\alpha$  then give the weighted average of the age-specific rates across the calendar periods, where the weights are the population sizes.

In simple terms this would give the model:

$\ln(R_{ij}) = \ln(\mu_{ij}) - \ln(N_{ij}) = \alpha_i = \frac{\sum_j (I_{ij} N_{ij})}{\sum_j N_{ij}}$
---

However, the previous analyses of the lung cancer incidence rates show that the age-specific male lung cancer rates are beginning to decline over calendar time (Figure 4.6 and 4.8) while the age-specific female lung cancer incidence rates are still increasing (Figure 4.7 and 4.9).

These increases may be due to either changes over calendar period or birth cohort. Models that incorporate



changes over period or changes over cohort are presented next.

If the age-specific rates plotted for each calendar period are parallel on the log scale, then it may be sensible to model the rates as:

$$\ln(R_{ij}) = \ln(\mu_{ij}) - \ln(N_{ij}) = \alpha_i + \beta_j$$

i.e. an additive effect of the age and period parameters in the log scale. For this model the design matrix  $X$  is given by

$$X = \{A_1, \dots, A_i, \dots, A_I, P_1, \dots, P_j, \dots, P_J\}$$

where  $A_i$  is the indicator vector for age group  $i$   
and  $P_j$  is the indicator vector for period  $j$

and the parameter vector is given by

$$\beta^T = \{\alpha_1, \dots, \alpha_i, \dots, \alpha_I, \beta_1, \dots, \beta_j, \dots, \beta_J\}$$

This may be an appropriate model for the male lung cancer incidence, where the age-period plots are approximately parallel on the log scale (Figure 4.6). Maximum likelihood methods may be used to fit this model, but the parameters will not be unique. If a constant ( $d$ ) is added to all the age parameters and the same constant is subtracted from all the period parameters, the likelihood would remain the same.

$$\begin{aligned}
l(\mu, I) &= \sum_{ij} (I_{ij} \ln(\alpha_i + \beta_j) - (\alpha_i + \beta_j)) \\
&= \sum_{ij} (I_{ij} \ln((\alpha_i+d) + (\beta_j-d)) - ((\alpha_i+d) + (\beta_j-d)))
\end{aligned}$$

However first order differences of the parameters would be unique

$$\begin{aligned}
\alpha_i - \alpha_x &= (\alpha_i+d) - (\alpha_x+d) \\
\text{and } \beta_j - \beta_y &= (\beta_j-d) - (\beta_y-d)
\end{aligned}$$

In this parameterisation, the first order differences between pairs of age parameters ( $\alpha_i - \alpha_x$ ) can be considered to be estimates of the log of the relative risk of age group  $i$  comparative to age group  $x$ . Thus for a given period  $j$

$$\begin{aligned}
[\alpha_i + \beta_j] - [\alpha_x + \beta_j] &= [\ln(\mu_{ij}) - \ln(N_{ij})] - [\ln(\mu_{xj}) - \ln(N_{xj})] \\
[\alpha_i - \alpha_x] &= \ln\{(\mu_{ij}/N_{ij}) / (\mu_{xj}/N_{xj})\}
\end{aligned}$$

which is the log of the relative risk of age group  $i$  comparative to age group  $x$  for all periods.

Similarly the first order differences between period parameters ( $\beta_j - \beta_y$ ) can be considered to be the log of the relative risk of period  $j$  compared to period  $y$ . Investigation of these differences allow deviations from linearity to be examined, and will be of importance in determining the relationship of cancer incidence to risk behaviours.

The effect of including calendar period in the model should be evaluated. For Poisson regression this can be determined by the change in the scaled deviance from the model including only the age parameters, to that also including the period parameters ( $S_{a,p;a}$ ) (Section 7.1). This change in scaled deviance is distributed as a  $X^2$  with  $J$  degrees of freedom.

The model described above is appropriate for the situation where the plot of the log rates by calendar period are parallel. If, however, the plot of the log rates by calendar period are not parallel, but the plots by birth cohort are, it is more sensible to include parameters for cohort effects in the model, and not period parameters. Thus the model would be:

$$\ln(R_{ik}) = \ln(\mu_{ik}) - \ln(N_{ik}) = \alpha_i + \gamma_k$$

The design matrix  $X$  would be given by

$$X = \{A_1, \dots, A_i, \dots, A_I, C_1, \dots, C_k, \dots, C_K\}$$

where  $A_i$  is the indicator vector for age group  $i$   
and  $C_k$  is the indicator vector for cohort  $k$ .

The vector of coefficients is given by:

$$\beta^T = \{\alpha_1, \dots, \alpha_i, \dots, \alpha_I, \gamma_1, \dots, \gamma_k, \dots, \gamma_K\}$$



This may be the appropriate model for female lung cancer, where the age-period plots are not parallel but the age-cohort plots are approximately parallel on the log scale (Figure 4.9). As with the age-period parameterisation, there is no unique solution and only first order differences can be uniquely estimated. The first order differences between parameters estimate the log relative risks between ages or cohorts.

As with the age-period models the effect of including cohort parameters in the model can be determined from the change in the scaled deviance ( $S_{a,c;a}$ ). The difference in this statistic between the model only including age effect and that also including cohort effects is distributed approximately as a  $X^2$  with  $K$  degrees of freedom.

In some instances both the age-period and the age-cohort models may seem appropriate; i.e. both the age-period and the age-cohort plots seem parallel. If this is the case, the differences between parameter estimates for both the adjacent periods and the adjacent cohorts will be approximately equal. This implies that the trends over both calendar period and birth cohort are linear. Clayton and Schiffers (1987a) recommend that this linearity is included in the model using a parameter for drift.

The log rates can be modelled as

$$\begin{aligned} \ln(R_{ij}) &= \ln(\mu_{ij}) - \ln(N_{ij}) = \alpha_i + \beta_j \\ &= \alpha_i + \delta_j (j - j_0) \end{aligned}$$

where  $\delta_j$  is the linear change between adjacent periods (or drift),  
and  $j_0$  is the reference period

It is possible to express the rate corresponding to  $R_{ij}$  as a rate indexed by age and cohort  $(i,k)$ . Thus

$$\ln(R_{ik}) = \ln(R_{ij}) = \alpha_i + \delta_j (k + i - I - (k_0 + i_0 - I))$$

which gives:

$$\ln(R_{ik}) = \alpha_i + \delta_j (i - i_0) + \delta_j (k - k_0)$$

In this situation it is not possible to determine whether the effect is a period or a cohort effect. Of note is the fact that for the two different models the first order differences of the age parameters are inflated by  $\delta_j$  per age interval if the period gradient constant is used in the age-cohort model.

Thus it is possible to use either model given above. If the age-period parameterisation is used, the age parameters will estimate the cross sectional age curve, while if the cohort parameterisation is used the age parameters will estimate the longitudinal age curve. In most cases there is no mechanism to distinguish between the two models.

The above section described the hierarchy of models that would be fitted to determine whether trends in cancer incidence are determined by either period or cohort effects. If both period and cohort effects contribute significantly to the fit, and neither the age-period model nor the age-cohort model fit the data adequately, then a more sophisticated model is required. The model will need to include age effects, period effects and cohort effects. The model that is fitted is:

$$\ln(R_{ij}) = \ln(\mu_{ij}) - \ln(N_{ij}) = \alpha_i + \beta_j + \gamma_k$$

Here the design matrix is:

$$X = \{A_1, \dots, A_I, P_1, \dots, P_J, C_1, \dots, C_K\}$$

where  $A_i$  is the indicator vector for age group  $i$   
 $P_j$  is the indicator vector for calendar period  $j$   
 $C_k$  is the indicator vector for birth cohort  $k$

From the relationship between the age, period and cohort indicators, viz:

$$k = I - i + j$$

the design matrix  $X$  can be shown to be singular (Rodgers 1982) and the maximum likelihood estimates of these parameters,  $\alpha$ ,  $\beta$ , and  $\gamma$  are not unique.

Thus the sets of parameter estimates can be considered to



$$\begin{aligned}
l(\mu, I) &= \sum_{ij} \{ I_{ij} \ln[\alpha_i + \beta_j + \gamma_k] - [\alpha_i + \beta_j + \gamma_k] \} \\
&= \sum_{ij} \{ I_{ij} \ln[(\alpha_i + di) + (\beta_j - dj) + (\gamma_k + d(k - I))] \\
&\quad - [(\alpha_i + di) + (\beta_j - dj) + (\gamma_k + d(k - I))] \}
\end{aligned}$$

be different rotations of each other (Figure 8.1a, 8.1b and 8.1c).

For these parameter estimates first order differences are also not unique:

$$\begin{aligned}
\beta_j^1 - \beta_y^1 &= (\beta_j - dj) - (\beta_y - dy) \\
&= (\beta_j - \beta_y) - (dj - dy)
\end{aligned}$$

This non-estimability of the parameters and their first order differences will not affect the projected values for the model.

If  $\beta_j^1 = \beta_j - dj$

then  $p^1 = p - d$

where  $p^1$  is the regression coefficient for the regression based on  $\beta^1$   
 $p$  is the regression coefficient for the regression based on  $\beta$   
and the intercepts  $a_b$  are the same for both regressions

Similarly if  $\gamma_j^1 = \gamma_j - d(k - I)$

then  $c^1 = c + d$

where  $c^1$  is the regression coefficient for the regression based on  $\gamma^1$   
 $c$  is the regression coefficient for the regression based on  $\gamma$   
and the intercepts  $a_c$  are the same for both regressions

Thus for any projected rate

$$\begin{aligned}
 \ln(R_{ij}) &= (\alpha_i + di) + (a_b + (p-d)j) + (a_c + (c+d)(k-I)) \\
 &= \alpha_i + (a + pj) + (a_c + c(k-I)) + di - dj + d(k-I) \\
 &= \alpha_i + (a + pj) + (a_c + c(k-I)) + d(i - j + k - I) \\
 &= \alpha_i + (a + pj) + (a_c + c(k-I))
 \end{aligned}$$

which is the projected value based on  $\alpha$ ,  $\beta$ , and  $\gamma$

Hence the final scaled deviance is also unaffected by the non-estimability of the parameters and their first order differences. By comparing the value of the final scaled deviance ( $S_{apc}$ ) to the values obtained from the age-period ( $S_{ap}$ ) and age-cohort ( $S_{ac}$ ) models separately, it is possible to determine the contributions of the period effects and the cohort effects to the fit (Table 8.1).

Second order differences are uniquely estimable:

$$\begin{aligned}
 &(\beta_{j-1}^1 - \beta_j^1) - (\beta_j^1 - \beta_{j+1}^1) \\
 &= [(\beta_{j-1} - d(j-1)) - (\beta_j - dj)] - [(\beta_j - dj) - (\beta_{j+1} - d(j+1))] \\
 &= [(\beta_{j-1} - \beta_j) - (\beta_j - \beta_{j-1})] - [d(j-1) - dj - dj + d(j+1)] \\
 &= [(\beta_{j-1} - \beta_j) - (\beta_j - \beta_{j-1})]
 \end{aligned}$$

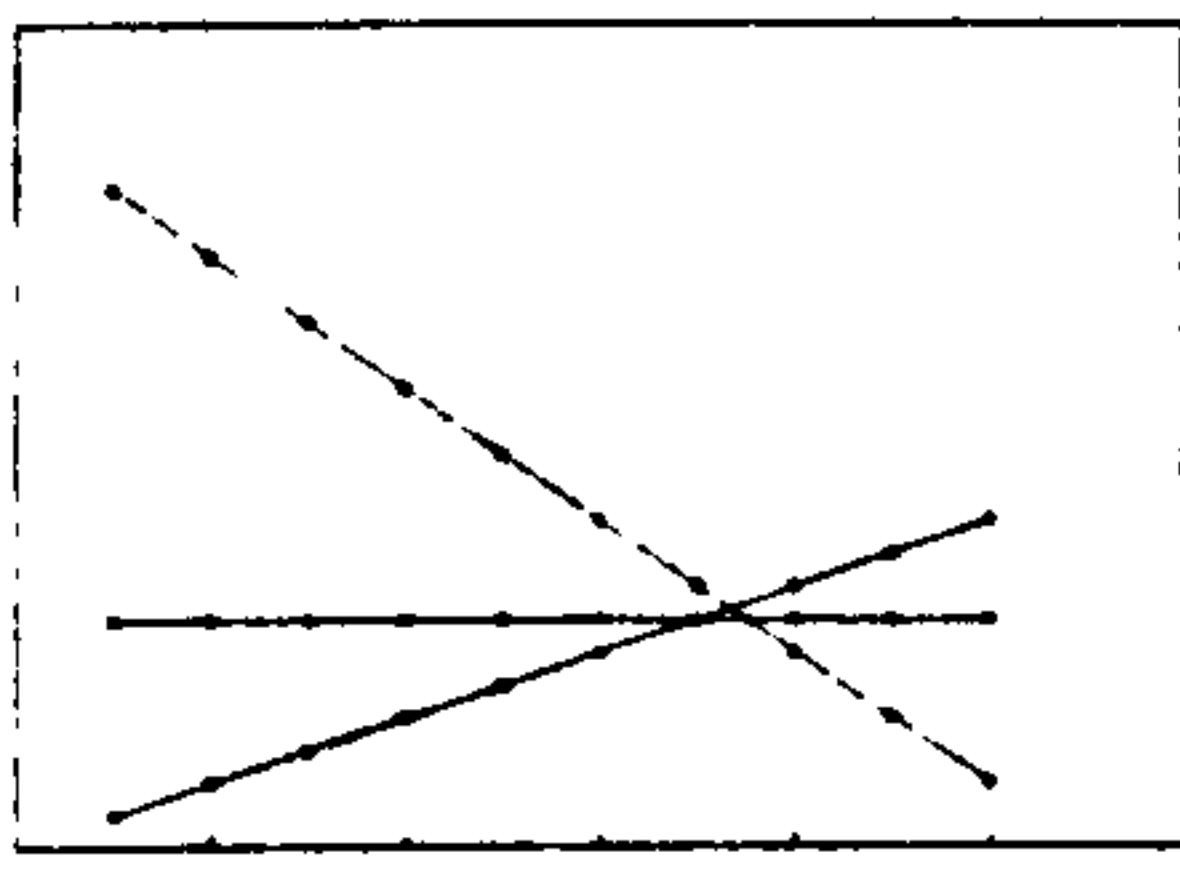
The identifiable second order differences can be interpreted as the change in slope of the trend curve around a central point. If one considers the difference:

$$\Delta_j = (\beta_{j-1} - \beta_j) - (\beta_j - \beta_{j+1})$$

which is the second order difference of the period estimates centred around period p then there are three possible interpretations for the different values of  $\Delta_j$ .

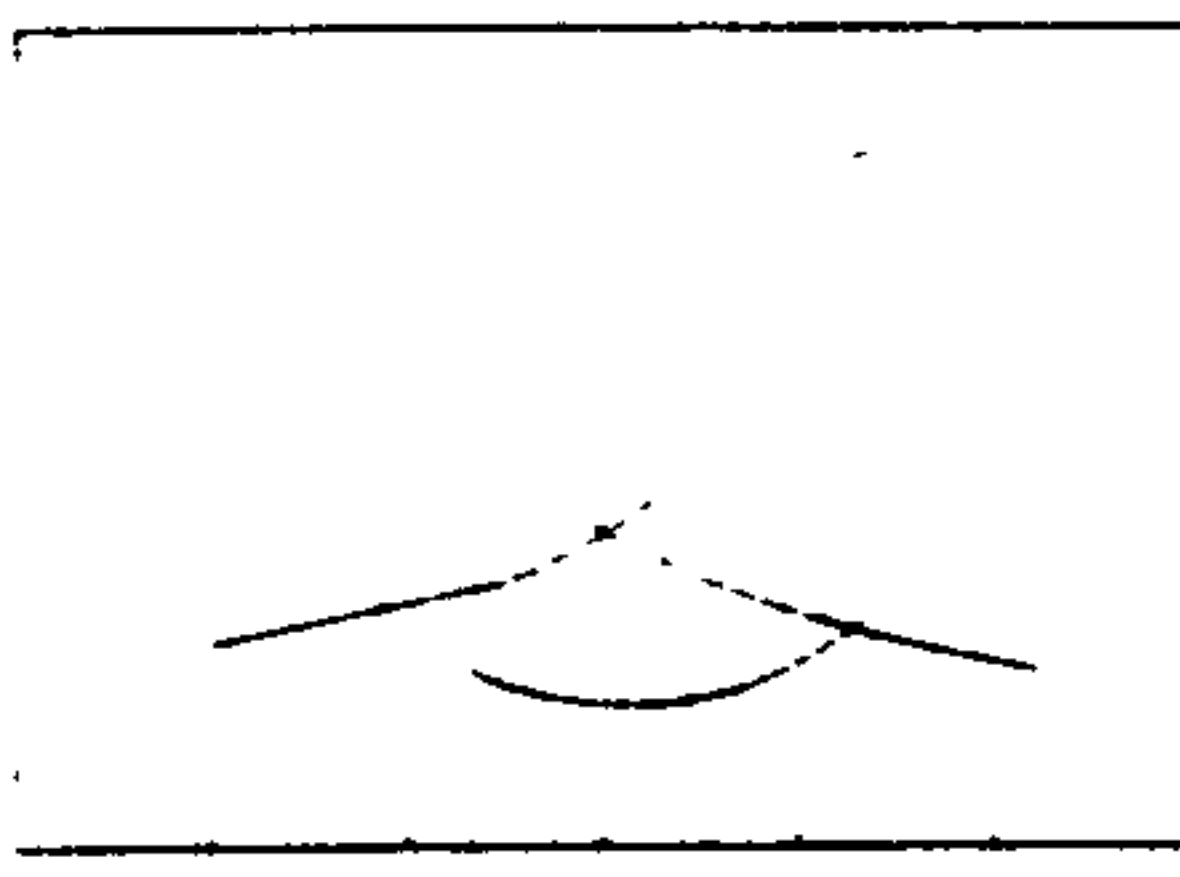
These are:

1.  $\Delta_j = 0$  the relationship between the logarithm of



the incidence rate and period is strictly linear between period (p-1) and period (p+1).

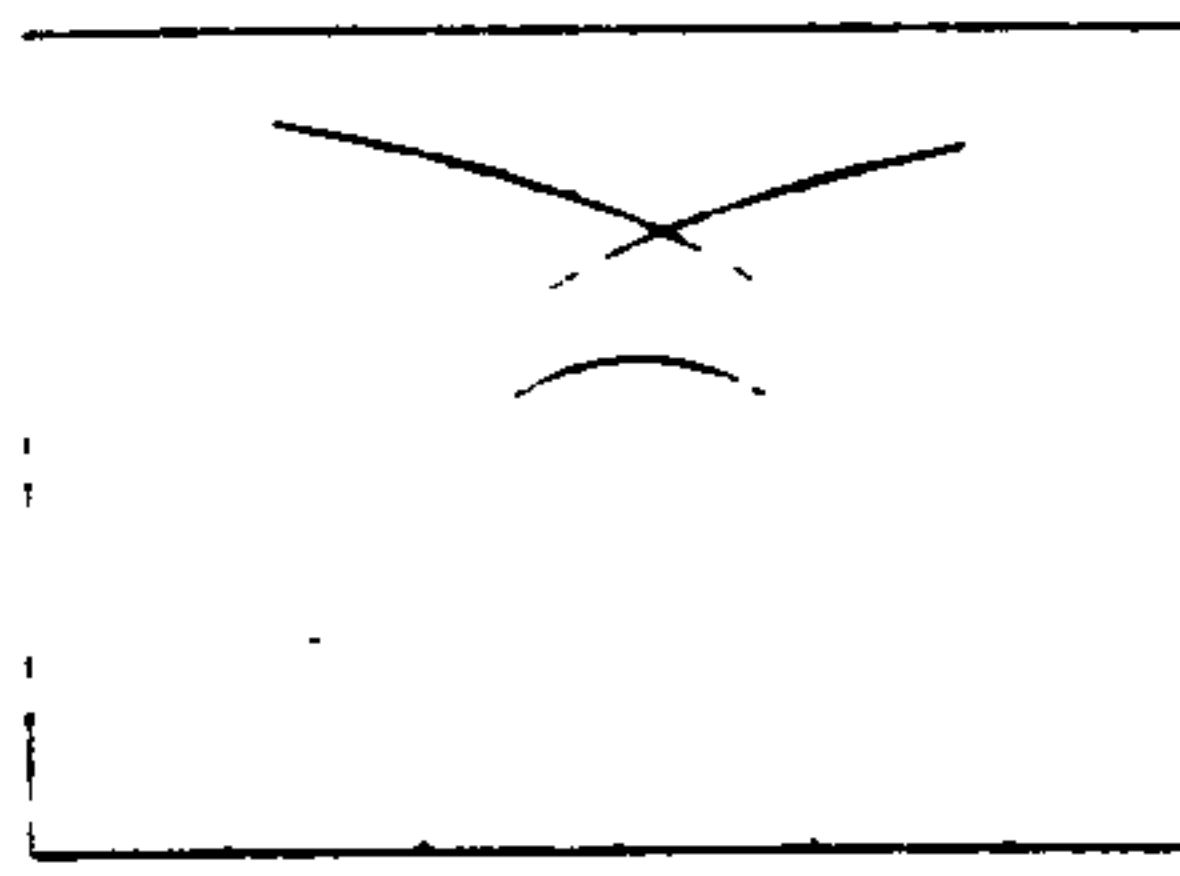
2.  $\Delta_j > 0$  the relationship between the logarithm of



the incidence rate and period could be interpreted as being strictly convex between period (p-1) and period (p+1), i.e. a line joining  $\beta_{j-1}$

and  $\beta_{j+1}$  will pass above the point  $\beta_j$ .

3.  $\Delta_j < 0$  the relationship between the logarithm of



the incidence rate and period could be interpreted as being strictly concave between period (p-1) and period group (p+1), i.e a line joining

$\beta_{j-1}$  and  $\beta_{j+1}$  will pass below the point  $\beta_j$ .

These properties hold true whatever the linear component of the trend. When interpreting these second order



differences it is of interest as to whether the curves are locally linear, convex or concave, and whether there are any changes in the type of curvature. Such changes would indicate possible underlying changes in the trends in risk behaviours, introduction of screening programs or other factors. For example female lung cancer incidence has been increasing linearly on the log scale ( $\Delta_j = 0$ ). If an effective anti smoking campaign are introduced, the expectation would be for rate of increase in female lung cancer incidence to slow. This would suggest that the relationship between lung cancer incidence and calendar period would be concave ( $\Delta_j < 0$ ).

#### 8.1.1 Fitting the Models

In order to determine whether the changes in incidence are due to period effects, cohort effects or periods and cohort effects a hierarchy of models are fitted using GLIM software (Baker and Nelder 1978). In this hierarchy the contribution of each set of effects is evaluated before investigating the next set (Table 8.1). In order for the age parameters ( $\alpha_i$ ) to estimate age-specific rates, the coefficient of the  $\ln(N_{ij})$  is constrained to be unity, and the intercept zero.

Table 8.1 Hierarchy of Model Fitting

<b>Model 1</b>	Age	
	$\alpha_i$	
	↓	
<b>Model 2</b>	Age + Drift	
	$\alpha_i + \delta_j(j-j_0)$	
	↙ ↘	
<b>Model 3a</b>	Age + Drift + Period	<b>Model 3b</b> Age + Drift + Cohort
	$\alpha_i + \delta_j(j-j_0) + \beta_j$	$\alpha_i + \delta_j(j-j_0) + \gamma_k$
	↙ ↘	↙
<b>Model 4</b>	Age + Period + Cohort	
	$\alpha_i + \beta_j + \gamma_k$	

The age parameters are included first in the model and evaluated for their contribution to the fit in the usual way. At this stage the model is:

<b>Model 1</b>	$\ln(R_{ij}) = \ln(\mu_{ij}) - \ln(N_{ij}) = \alpha_i$
----------------	--

Then the contribution of the linear drift parameter is assessed. The model now contains a parameter corresponding to linear drift over the periods and the model becomes:

<b>Model 2</b>	$\ln(R_{ij}) = \ln(\mu_{ij}) - \ln(N_{ij}) = \alpha_i + \delta_j(j - j_0)$
----------------	--

This model may account adequately for the variation in lung cancer incidence rates over time and therefore be used as the means of projection. If this model does not fit the data adequately, the additional contribution of

the non linear period effects are then assessed. The model becomes:

$$\begin{aligned} \text{Model 3a } \ln(R_{ij}) &= \ln(\mu_{ij}) - \ln(N_{ij}) \\ &= \alpha_i + \delta_j (j - j_0) + \beta_j \end{aligned}$$

In this model the linear effects are accounted for by the  $\delta_j$  and the  $\beta_j$  represents the non linear effects for calendar period. The contribution of these non linear period effects is determined by the change in the scaled deviance between Model 3a and Model 2 ( $S_{a,\delta,p;a,\delta}$ ) and the distribution of the residuals.

As discussed earlier, the changes in the age-specific incidence rates may be due to non-linear cohort effects (Section 7.2.3). The addition of these effects to the drift model are also evaluated.

For this the model is:

$$\begin{aligned} \text{Model 3b } \ln(R_{ij}) &= \ln(\mu_{ij}) - \ln(N_{ij}) \\ &= \alpha_i + \delta (j - j_0) + \gamma_k \end{aligned}$$

The parameters  $\gamma_k$  represent the non linear cohort effects.



If the period parameters are the only significant contribution to the fit of the model, and that fit is adequate, then the trends over time are due to calendar period effects. In this case a final age-period model is fitted, viz

$$\text{Model 3a' } \ln(R_{ij}) = \ln(\mu_{ij}) - \ln(N_{ij}) = \alpha_i + \beta_j$$

The parameters  $\beta_j$  include both linear and nonlinear components. Examination of both age and period parameters would allow investigation of the shape of the age and period curves, but not the unique values of the age or period parameters.

Similarly if the cohort parameters are the only significant contribution to the model, then it can be assumed that the trends over time are due to cohort effects. In this case a final age-cohort model is fitted, viz:

$$\text{Model 3b' } \ln(R_{ij}) = \ln(\mu_{ij}) - \ln(N_{ij}) = \alpha_i + \gamma_k$$

As for the age-period model, examination of the age and cohort parameters would allow the investigation of the form of the age and cohort curves, but not the values of the age or cohort parameters.

If both period effects and cohort effects contribute significantly to the model and neither model 3a' nor 3b' fit adequately, then the age-period-cohort model is fitted to the data, viz:

$$\text{Model 4 } \ln(R_{ij}) = \ln(\mu_{ij}) - \ln(N_{ij}) = \alpha_i + \beta_j + \gamma_k$$

The additional contribution of the cohort effects to the model containing age and period effects (Model 3a') is assessed. The additional contribution of the period effects to the model containing cohort effects (Model 3b') is also assessed. If both periods and cohort contribute significantly to the smaller models, it is assumed that the variation in the incidence rates is due to age and period and cohort effects.

## 8.2 Results

This section will present the results from the investigations of the trends in cancer incidence over age, calendar period and birth cohort. The results will be presented separately for each sex according to the hierarchy of models presented in Table 8.1.

That is:

Model 1	Age
Model 2	Age + Drift
Model 3a	Age + Drift + Period
Model 3b	Age + Drift + Cohort
Model 4	Age + Period + Cohort

### 8.2.1 Male Lung Cancer

Five year age-specific frequencies for male lung cancer, and the corresponding population figures are used in these analyses. The age range is restricted to 30 to 84 for the reasons outlined in Section 8.1, and the calendar periods used are those quinquennia centred on 1961, 1966, 1971, 1976, 1981 and 1986. Thus the data include 11 five year age groups and 6 five year calendar periods (Table 8.2).

The first model fitted is the age model (Model 1). This model includes only age effects. The change in scaled deviance from the model with no parameters ( $S_{a,0} = 51505$ ,  $df = 11$ ) indicates that age does contribute to the model, but the final scaled deviance ( $S_a = 828.92$ ,  $df = 55$ ) indicates that the fit is not adequate.



**Table 8.2a Male Lung Cancer Incidence Frequencies,  
LRHB 1959-1973, MRHA 1974-1988**

Age Group	Quinquennia					
	1961	1966	1971	1976	1981	1986
30-34	25	15	16	14	11	6
35-39	43	44	37	32	35	31
40-44	125	120	122	83	69	60
45-49	279	276	303	285	208	152
50-54	561	507	490	592	484	330
55-59	940	865	901	946	853	699
60-64	1097	1278	1390	1413	1224	1192
65-69	975	1291	1517	1697	1436	1465
70-74	737	970	1150	1486	1650	1432
75-79	385	577	624	973	1143	1160
80-84	135	234	273	419	549	641
<b>Total</b>	<b>5302</b>	<b>6177</b>	<b>6823</b>	<b>7940</b>	<b>7662</b>	<b>7168</b>

**Table 8.2b Male Population Frequencies (1,000s)  
LRHB 1959-1973, MRHA 1974-1988**

Age Group	Quinquennia					
	1961	1966	1971	1976	1981	1986
30-34	355.5	338.0	328.5	385.5	442.0	398.0
35-39	361.0	321.5	324.0	356.5	377.5	433.5
40-44	332.0	341.0	339.0	358.0	351.0	377.5
45-49	351.5	324.0	341.5	370.5	346.5	339.0
50-54	347.5	326.0	307.5	372.0	356.5	331.5
55-59	306.5	311.0	311.5	330.0	345.0	327.5
60-64	235.5	264.0	281.5	313.0	291.0	309.5
65-69	173.5	194.5	219.0	263.5	260.5	247.0
70-74	122.0	126.5	138.0	188.0	201.5	202.0
75-79	75.0	77.5	79.5	100.5	123.0	137.0
80-84	37.5	35.0	41.0	46.5	53.0	67.0
<b>Total</b>	<b>2697.5</b>	<b>2659.0</b>	<b>2711.0</b>	<b>3084.0</b>	<b>3147.5</b>	<b>3169.5</b>

The model including the parameter for drift is fitted next. Linear drift over periods ( $\delta_j$ ) contributes significantly to the model ( $S_{a\delta;a} = 13.92$ ,  $df = 1$ ), and indicates that incidence is declining over the period of investigation ( $\delta_j = -0.0109$ ). Even so, the final fit is not adequate ( $S_{a\delta}=814.99$ ,  $df=54$ ) This suggests that other factors, such as non-linear period effects and/or cohort effects, are contributing to the trends in cancer incidence. Thus the addition of non-linear calendar period and birth cohort parameters are investigated.

The addition of the non linear calendar parameters (Model 3a) to the age + linear drift model improved the fit ( $S_{a\delta p;a\delta} = 155.1$ ,  $df = 4$ ), but there is still excess variation ( $S_{a\delta p}=659.88$ ,  $df=50$ ).

The addition of the non-linear cohort parameters (Model 3b) to the age + drift model improves the fit markedly ( $S_{a\delta c;a\delta} = 731.3$ ,  $df = 14$ ), but the final scaled deviance still indicates a lack of fit of the data ( $S_{a\delta c}=83.70$ ,  $df=40$ ).

Given the significant lack of fit of the models including age, drift and period, and age, drift and cohort (Models 3a and 3b), the full age-period and cohort model (Model 4) is now fitted.

Inclusion of the period effects in the age-cohort model

contributed significantly to the fit ( $S_{a\delta cp, a\delta c}=25.83$ ,  $df=4$ ). Similarly addition of the cohort effects contributed significantly to the age-period model ( $S_{a\delta pc, a\delta p}=602.01$ ,  $df=14$ ). The final scaled deviance is still significant ( $S_{a\delta pc} = S_{a\delta cp} = 57.867$ ,  $df = 36$ ). Examination of the residuals showed no consistent pattern and therefore this lack of fit may be due to over dispersion. Thus the age-period-cohort model is chosen as the 'best' model for fitting the data.

The parameter estimates from the final fit are given in Table 8.3, but these estimates are not unique (Section 8.2). The second order differences are estimable (Table 8.4) and show that the period parameters form a concave curve, and the form of the cohort curve is also generally concave. The exceptions are the cohorts born in 1909-1918 and 1929-1939 and 1934-1943. However, the magnitude of the period second order differences are smaller than those for cohort or age.

Two sets of parameter values are plotted to illustrate the variations in their values with different constraints (Set 1  $\beta_1 = \beta_J = 0$ ; Set 2  $\gamma_1 = \gamma_K = 0$ ). The two sets of parameter values are rotations of each other (Figures 8.1a, 8.1b and 8.1c).



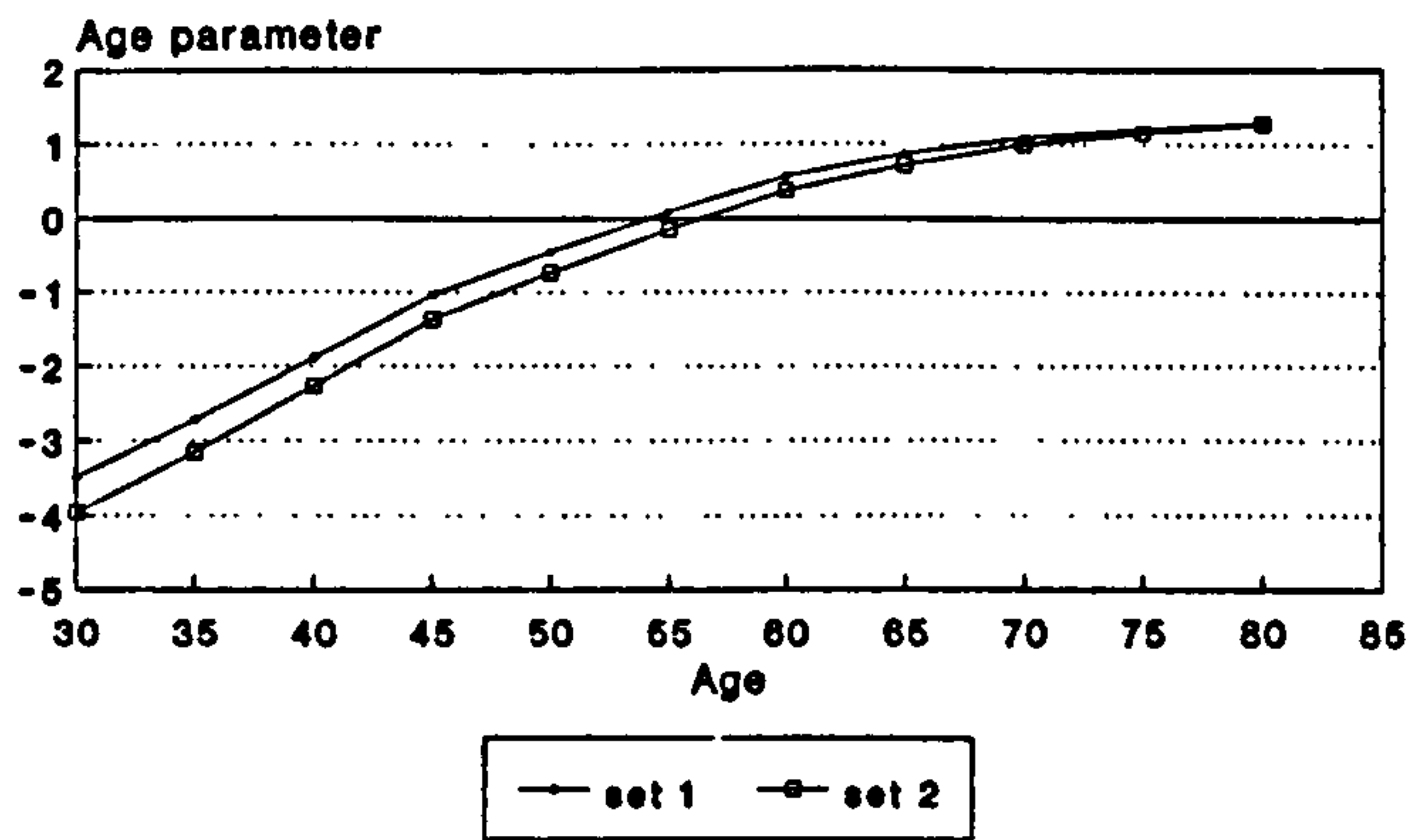
Table 8.3 Parameter Estimates for Male Lung Cancer from the Age Period Cohort Modelling

Age	Parameter Estimate	Cohort	Parameter Estimate
30-34	-3.958	1874-83	0.000
35-39	-3.151	1879-88	0.530
40-44	-2.265	1884-93	0.790
45-49	-1.370	1889-98	0.999
50-54	-0.730	1894-1903	1.175
55-59	-0.142	1899-1908	1.214
60-64	0.371	1904-13	1.228
65-69	0.724	1909-18	1.189
70-74	0.997	1914-23	1.258
75-79	1.156	1919-28	1.227
80-84	1.281	1924-33	1.176
Period		1929-38	0.970
1959-63	0.000	1934-43	0.814
1964-68	-0.001	1939-48	0.750
1969-73	-0.026	1944-53	0.654
1974-78	-0.083	1949-58	0.000
1979-83	-0.160		
1984-88	-0.237		

Table 8.4 Second Order Differences in Parameter Estimates for Male Lung Cancer from the Age-Period-Cohort Modelling

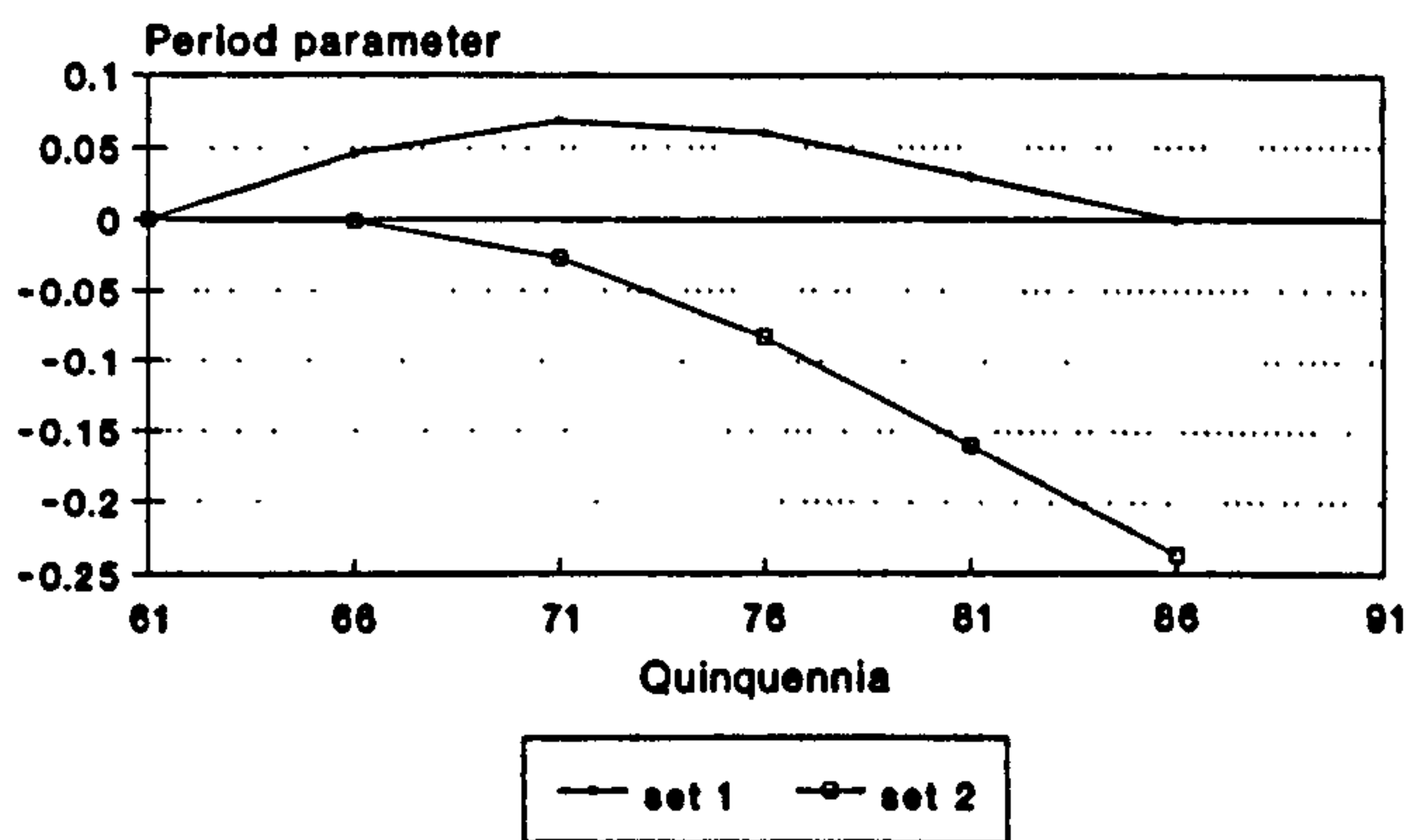
Age	Parameter Difference	Cohort	Parameter Difference
35-39	0.079	1879-88	-0.271
40-44	0.009	1884-93	-0.051
45-49	-0.255	1889-98	-0.033
50-54	-0.051	1894-1903	-0.137
55-59	-0.075	1899-1908	-0.025
60-64	-0.160	1904-13	-0.053
65-69	-0.081	1909-18	0.108
70-74	-0.114	1914-23	-0.100
75-79	-0.034	1919-28	-0.020
		1924-33	-0.155
Period		1929-38	0.506
		1934-43	0.092
1964-68	-0.024	1939-48	-0.032
1969-73	-0.031	1944-53	-0.558
1974-78	-0.021	1949-58	0.000
1979-83	-0.001		

Figure 8.1a Age Parameter Estimates  
Male Lung Cancer, Age-Period-Cohort  
Analysis, 1959-1988



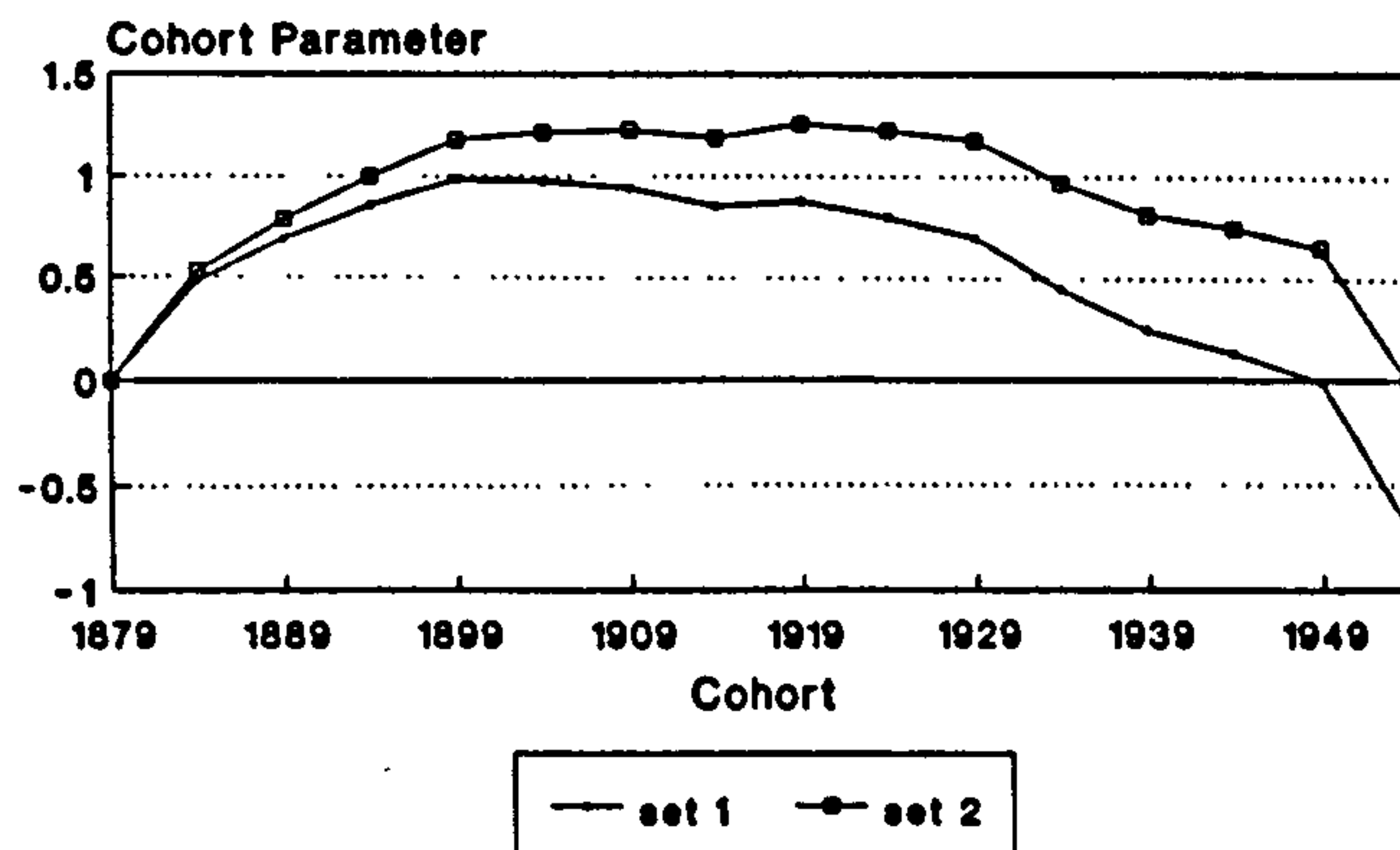
Mersey Regional Cancer Registry

Figure 8.1b Period Parameter Estimates  
Male Lung Cancer, Age-Period-Cohort  
Analysis, 1959-1988



Mersey Regional Cancer Registry

Figure 8.1c Cohort Parameter Estimates  
Male Lung Cancer, Age-Period-Cohort  
Analysis, 1959-1988



Mersey Regional Cancer Registry



### 8.2.2 Female Lung Cancer

Five year age-specific frequencies for female lung cancer, and the corresponding population figures are used in these analyses. The age range is restricted to 35 to 84 years for the reasons outlined in Section 8.1, and the calendar period used are the same as those quinquennia used for male lung cancer. Thus the data covered 10 five year age groups and 6 five year calendar periods (Table 8.5)

This model including only the age effects is fitted first. The parameter estimates for all age groups are lower than those for the males. The change in scaled deviance from the model with no parameters ( $S_{a,0} = 7269$ ,  $df = 11$ ) indicates that age does contribute to the model, but the final scaled deviance ( $S_a = 1698.6$ ,  $df = 50$ ) indicates that the fit is not adequate. This is to be expected as the incidence rates have been increasing over the period of investigation.

The model including the parameter for linear drift is fitted next. Linear drift over periods ( $\delta_j$ ) contributes significantly to the model ( $S_{a\delta,a} = 1435.$ ,  $df=1$ ), and the sign of the coefficient confirms that incidence is increasing over the period of investigation ( $\delta_j = 0.2124$ ) As with the males, the final fit is not adequate ( $S_{a\delta} = 263.40$ ,  $df=49$ ), and other effects are investigated.

**Table 8.5a Female Lung Cancer Incidence Frequencies,  
LRHB 1959-1973, MRHA 1974-1988**

Age Group	Quinquennia					
	1961	1966	1971	1976	1981	1986
35-39	18	18	20	12	19	13
40-44	38	45	41	41	38	36
45-49	65	68	85	112	75	89
50-54	101	137	166	214	184	155
55-59	120	169	214	312	373	349
60-64	143	161	259	433	481	621
65-69	135	189	290	412	527	622
70-74	117	180	237	362	507	605
75-79	100	124	169	274	342	461
80-84	60	94	100	122	218	284
<b>Total</b>	<b>897</b>	<b>1185</b>	<b>1581</b>	<b>2294</b>	<b>2764</b>	<b>3235</b>

**Table 8.5b Female Population Frequencies (1,000s)  
LRHB 1959-1973, MRHA 1974-1988**

Age Group	Quinquennia					
	1961	1966	1971	1976	1981	1986
30-34	375.0	327.5	315.5	356.5	375.5	435.0
35-39	357.0	360.0	337.0	355.5	355.0	377.0
40-44	379.5	348.5	355.0	375.5	346.5	339.5
45-49	372.5	359.5	333.0	389.5	364.5	333.0
50-54	339.5	350.5	344.0	365.0	369.5	343.0
55-59	302.0	322.0	328.0	367.0	339.0	345.0
60-64	254.5	281.0	285.5	337.5	334.0	306.5
65-69	204.5	225.5	230.5	285.0	294.0	286.5
70-74	144.5	160.5	163.0	208.5	227.0	237.5
75-79	84.0	96.5	101.0	124.0	142.5	159.5
80-84	37.5	35.0	41.0	46.5	53.0	67.0
<b>Total</b>	<b>2850.5</b>	<b>2866.5</b>	<b>2833.5</b>	<b>3210.5</b>	<b>3200.5</b>	<b>3229.5</b>

The addition of the non linear calendar parameters (Model 3a) to the age + linear drift model improves the fit ( $S_{a\delta p, a\delta} = 11.836$ ,  $df = 4$ ), but there is still excess variation ( $S_{a\delta p} = 251.56$ ,  $df = 45$ ).

The addition of the non-linear cohort parameters (Model 3b) to the age + drift model improves the fit markedly ( $S_{a\delta c, a\delta} = 220.6$ ,  $df = 13$ ). The final scaled deviance indicates a good fit of the data ( $S_{a\delta c} = 42.83$ ,  $df = 36$ ). Examination of the residuals show no distinct patterns.

The age-period-cohort model is then fitted to confirm that the addition of the period effects to the age-cohort model does not significantly improve the fit. Inclusion of the period effects in the age-cohort model did not contribute significantly to the fit ( $S_{a\delta cp, a\delta c} = 6.54$ ,  $df = 4$ ). Thus the age-cohort model is the 'best' model for the data. This model does not include the drift parameter, since the cohort parameters would include linear and non linear effects.

The parameter estimates from the final fit are given in Table 8.6, but as described in section 8.1 these estimates are not unique. The first order differences are estimable (Table 8.7) Two sets of parameter values are plotted (Figures 8.2a and 8.2b) to illustrate the effect of adding a constant to the age parameters and subtracting the same constant from the cohort parameters



Table 8.6 Parameter Estimates for Female Lung Cancer from the Age Cohort Modelling, LRHB 1959-1973, MRHA 1974-1988

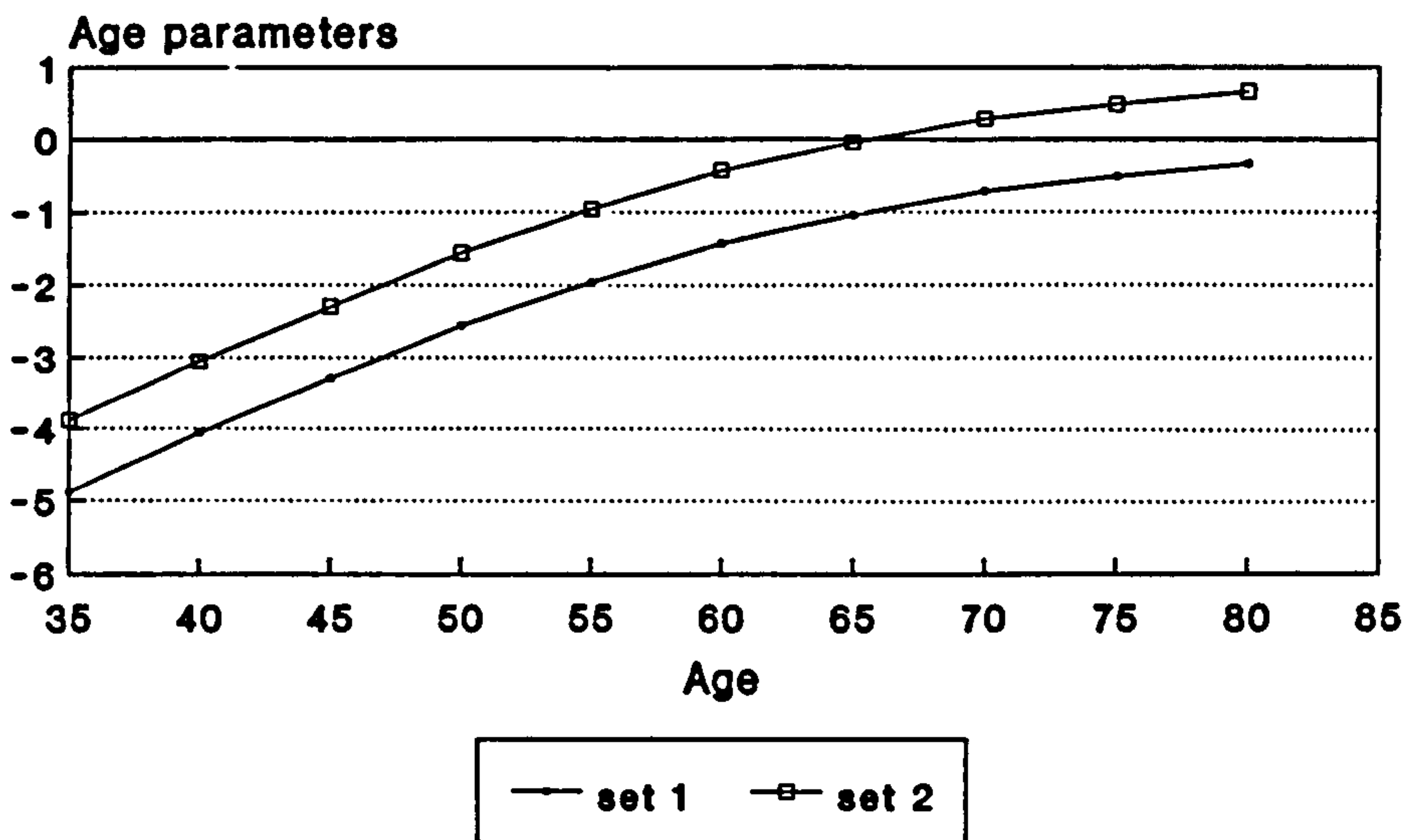
Age	Parameter Estimate	Cohort	Parameter Estimate
35-39	-4.886	1874-83	0.000
40-44	-4.058	1879-88	0.222
45-49	-3.298	1884-93	0.242
50-54	-2.557	1889-98	0.453
55-59	-1.958	1894-1903	0.732
60-64	-1.426	1899-1908	0.931
65-69	-1.045	1904-13	1.224
70-74	-0.719	1909-18	1.514
75-79	-0.514	1914-23	1.776
80-84	-0.337	1919-28	1.979
		1924-33	1.963
		1929-38	1.820
		1934-43	1.873
		1939-48	1.772
		1944-53	1.376

(Set 1:  $\{\alpha, \gamma\}$ ; Set 2:  $\{(\alpha+1), (\gamma-1)\}$ ). As would be expected the age parameter coefficients increase with increasing age, in an S shaped curve. The cohort coefficients indicate that the cohort effects increase until the cohort born in 1924 to 1933, and then begin to decline slowly.

Table 8.7 First Order Differences in Parameter Estimates for Female Lung Cancer from the Age Cohort Modelling

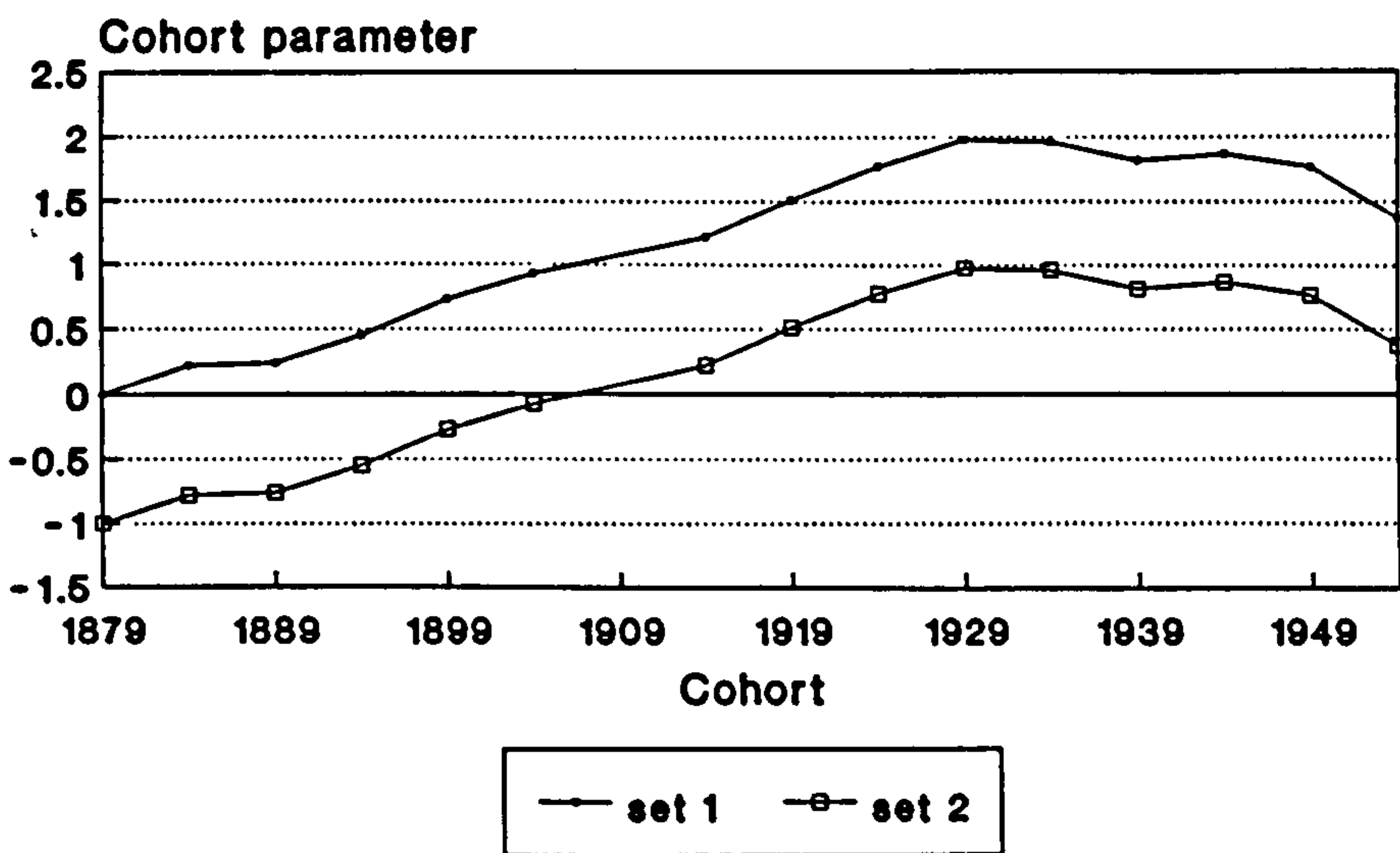
Age	Parameter Difference	Cohort	Parameter Difference
40-44	0.828	1879-88	0.222
45-49	0.760	1884-93	0.019
50-54	0.720	1889-98	0.211
55-59	0.599	1894-1903	0.280
60-64	0.532	1899-1908	0.199
65-69	0.381	1904-13	0.293
70-74	0.326	1909-18	0.290
75-79	0.204	1914-23	0.262
80-84	0.178	1919-28	0.203
		1924-33	-0.160
		1929-38	-0.143
		1934-43	0.053
		1939-48	-0.101
		1944-53	-0.396
		1949-58	0.000

**Figure 8.2a Age Parameter Estimates  
Female Lung Cancer, Age-Cohort Analysis  
1959-1988**



Mersey Regional Cancer Registry

**Figure 8.2b Cohort Parameter Estimates  
Female Lung Cancer, Age-Cohort Analysis  
1959-1988**



Mersey Regional Cancer Registry



### 8.2.3 Summary of Model Fitting

These results have shown that the age-specific rates for both male and female lung cancer are influenced by birth cohort. In both sexes the age-specific incidence is decreasing for the more recent cohorts (Figures 4.8 and 4.9). These decreases are confirmed by the results from the age-period-cohort analyses. The trends in male lung cancer incidence are also influenced by changes over the calendar period. This is not the case for females.

The results also help to explain the lack of fit for the age standardised rates and the broad-band age-specific rates. In both of these analyses, the trends in smoking behaviours are fitted to changes over calendar period. For the males the trends in tar content of cigarettes is likely to be a period effect, but both the quantity smoked and the percentage of the population smoking are likely to be cohort effects. In the analyses of the age-standardised rates and the broad-band age-specific rates, the smoking factors that best explained the trends in lung cancer incidence are the number of cigarettes smoked per person or per smoker and the percentage of the population who smoke. These are the best predictors for both males and females. The average tar content of cigarettes is an additional significant predictor for males, but not for females. This agrees well with the results of the age-period-cohort analyses. For males both

period and cohort variation were significant, while for females the trends could be explained by cohort variation alone.

### 8.3 Projections

#### 8.3.1 Methods

The age-period-cohort models determined for male and female lung cancer incidence can be used for estimating future lung cancer incidence. For males the final model included parameters for age, period and cohort effects.

$$\ln(R_{ij}) = \ln(\mu_{ij}) - \ln(N_{ij}) = \alpha_i + \beta_j + \gamma_k$$

For females only parameters for age and cohort are included in the final model.

$$\ln(R_{ij}) = \ln(\mu_{ij}) - \ln(N_{ij}) = \alpha_i + \gamma_k$$

It is assumed that the age parameters ( $\alpha_i$ ) remain constant over the period of projection. For males estimates for the period parameters ( $\beta_j$ ) for the quinquennia centred on 1961, 1966, 1971, 1976, 1981 and 1986 are available from the analysis. Estimates of the period parameters for 1991, 2001 and 2011 are necessary for projections for these years.

Osmond (1985) proposed extrapolating the period parameter

curve, using simple linear regression, to estimate the future period parameters. This assumes that the trend over the more recent calendar periods is linear. The number of past periods to include in the regressions would be determined by visual inspection of the period parameter curve. The period parameter curve for the whole period 1961 to 1986 is non-linear (Figure 8.1b). Only the parameter estimates for 1971, 1976, 1981 and 1986, where the curve appears linear, are included in this regression.

The final models of lung cancer incidence for both males and females include cohort effects. Estimates of cohort parameters are available from the age-period-cohort analyses for the cohorts 1874-83 through 1949-58 for males and 1874-83 through 1944-53 for females. For projections for 1991 it is necessary to estimate the cohort parameter for the birth cohort in 1954-63 for the males and 1949-58 for the females (Table 8.8). For the projections for 2001 it is also necessary to estimate the cohort parameters for those birth cohorts 1959-68 and 1964-73 for males, and for females those birth cohorts for 1954-63 and 1959-68. For the projections for 2011 birth cohort parameters need to be estimated for the 1969-78 and 1974-83 birth cohorts for males and the 1964-73 and 1969-78 birth cohorts for females.



**Table 8.8 Age Groups, Periods and Cohorts**

Age	Period									
	1959-63	1964-68	1969-73	1974-78	1979-83	1984-88	1989-93	1999-03	2009-13	
30-34	1924-33	1929-38	1934-43	1939-48	1944-53	1949-58	1954-63	1964-73	1974-83	
35-39	1919-28	1924-33	1929-38	1934-43	1939-48	1944-53	1949-58	1959-68	1969-78	
40-44	1914-23	1919-28	1924-33	1929-38	1934-43	1939-48	1944-53	1954-63	1964-73	
45-49	1909-18	1914-23	1919-28	1924-33	1929-38	1934-43	1939-48	1949-58	1959-68	
50-54	1904-13	1909-18	1914-23	1919-28	1924-33	1929-38	1934-43	1944-53	1954-63	
55-59	1899-08	1904-13	1909-18	1914-23	1919-28	1924-33	1929-38	1939-48	1949-58	
60-64	1894-03	1899-08	1904-13	1909-18	1914-23	1919-28	1924-33	1934-43	1944-53	
65-69	1889-98	1894-03	1899-08	1904-13	1909-18	1914-23	1919-28	1929-38	1939-48	
70-74	1884-93	1889-98	1894-03	1899-08	1904-13	1909-18	1914-23	1924-33	1934-43	
75-79	1879-88	1884-93	1889-98	1894-03	1899-08	1904-13	1909-18	1919-28	1929-38	
80-84	1874-83	1879-88	1884-93	1889-98	1894-03	1899-08	1904-13	1914-23	1924-33	

NB *Italics* indicate those cohort parameters that have to be estimated for the projections  
The 1949-58 cohort parameters only need to be estimated for females

Osmond (1985) recommends the linear extrapolation of the recent linear section of the cohort parameter curve. The parameter estimate for the most recent cohort (birth-cohort 1949-58 for males and 1944-53 for females) is excluded from these calculations. The estimates for these birth-cohorts are likely to be imprecise because they are based on only one observation, and that is for the youngest age group where the number of incident cancer cases is low. This cohort parameter estimate does not contribute to the final scaled deviance, since the estimate ensures that the model fits this point exactly. The number of data points to include in the regressions is determined by visual examination of the data, and in these analyses the 6 most recent cohorts (excluding the last one) are included.

The age specific rates estimated using linear extrapolation of the period and/or cohort parameter estimates can be shown to be unique (Osmond 1985).

The Health of the Nation targets require a 30% reduction in lung cancer for males, and a 15% reduction for females, between the years 1990 and 2010. In order to assess whether the Health of the Nation targets will be met for Mersey Region, the age-standardised rates for the age groups 30-74 for males and 35-74 for females are calculated from the projected age-specific rates, for the years 1991 and 2011 using the Mersey region population

for 1991 as the standard. Data for 1991 and 2011 are used, rather than 1990 and 2010, because the analysis uses five year lung cancer incidence rates, centred on the census years. The population for the year 1991 is used as the standard because 1991 is the year against which comparisons are to be made, and 1991 is also the most recent census year. The Mersey Region population is used as the standard, rather than the standard European population, because the comparison is to be made within the region and therefore this population is more appropriate. The projected percentage change in lung cancer incidence is calculated for both males and females for comparison with the Health of the Nation targets.

The age-specific frequencies of lung cancer incidence are used to determine service needs for the years 2001. These are calculated using the age-specific treatment pattern for each sex recorded for 1983-87 (Williams et al 1993).

### **8.3.2 Results**

The estimates for the calendar period parameters and the birth cohort parameters for male lung cancer are decreasing over time (Table 8.9). The estimates for the cohort parameters for females are also decreasing (Table 8.9)



Table 8.9 Estimated Period and Cohort Parameters for Lung Cancer, MRHA

Period	Parameter	Cohort	Parameter	
	Male		Male	Female
1991	-0.3036	1944-53		1.730
2001	-0.4453	1949-58	0.4933	1.680
2011	-0.5869	1954-63	0.3669	1.629
		1959-69	0.2405	1.579
		1964-73	0.1141	1.529
		1969-78	-0.0123	1.479
		1974-83	-0.1387	

For males the projected age-specific incidence rates are calculated from the equation:

$$\ln(R_{apc}) = \alpha_a + \beta_p + \gamma_c$$

The age-period-cohort analyses have used rates per 1,000. In order to determine the incidence rates per 100,000 the estimates of the lung cancer incidence rates are multiplied by 100.

For example for the youngest age group in 1991 (30-34 years) the log of the rate is given by:

$$\begin{aligned}\ln(R) &= \alpha_{30-34} + \beta_{1991} + \gamma_{1954-63} \\ &= -3.958 - 0.3036 + 0.3669 \\ &= -3.8947 \\ R &= \exp(-3.8947) \\ &= 0.0203 \text{ cases/1,000} \\ &= 2.03 \text{ cases/100,000}\end{aligned}$$

Thus the projected age-specific rate for the 30-34 year age group in 1991 is 0.0203 per 1,000 or 2.03 per 100,000. It should be noted that in 1991 the cohort parameters for the age groups 40-44 years and above are those that are estimated in the main analysis (Table 8.8).

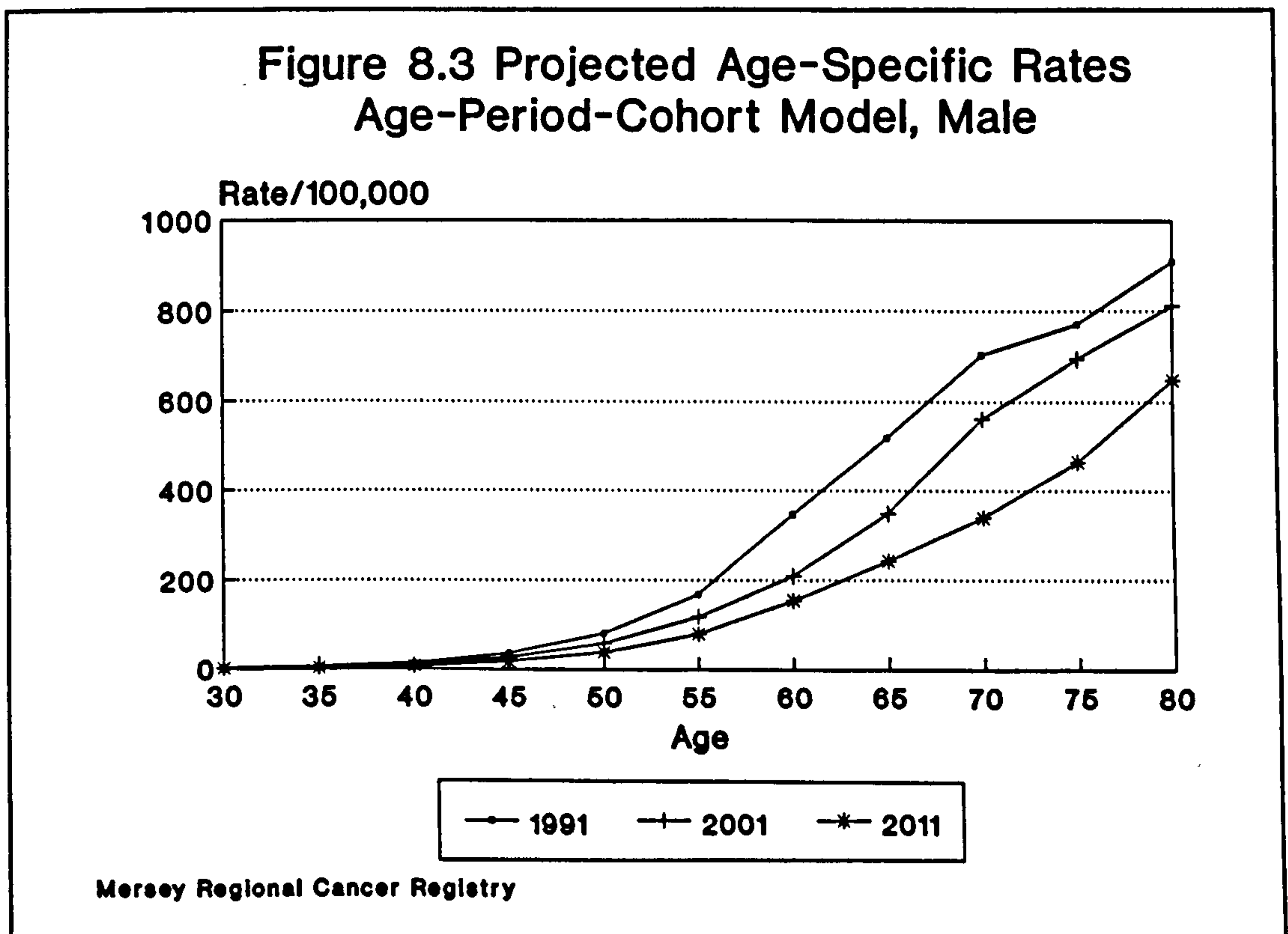
For females the projected age-specific incidence rates are calculated from the equation:

$$\ln(R_{apc}) = \alpha_a + \gamma_c$$

For example for the youngest age group (35-39 years) in 1991 the log of the rate is given by:

$$\begin{aligned}
 \ln(R) &= \alpha_{35-39} + \gamma_{1949-58} \\
 &= -4.886 + 1.680 \\
 &= -3.206 \\
 R &= \exp(-3.206) \\
 &= 0.0405 \text{ cases/1,000} \\
 &= 4.05 \text{ cases/100,000}
 \end{aligned}$$

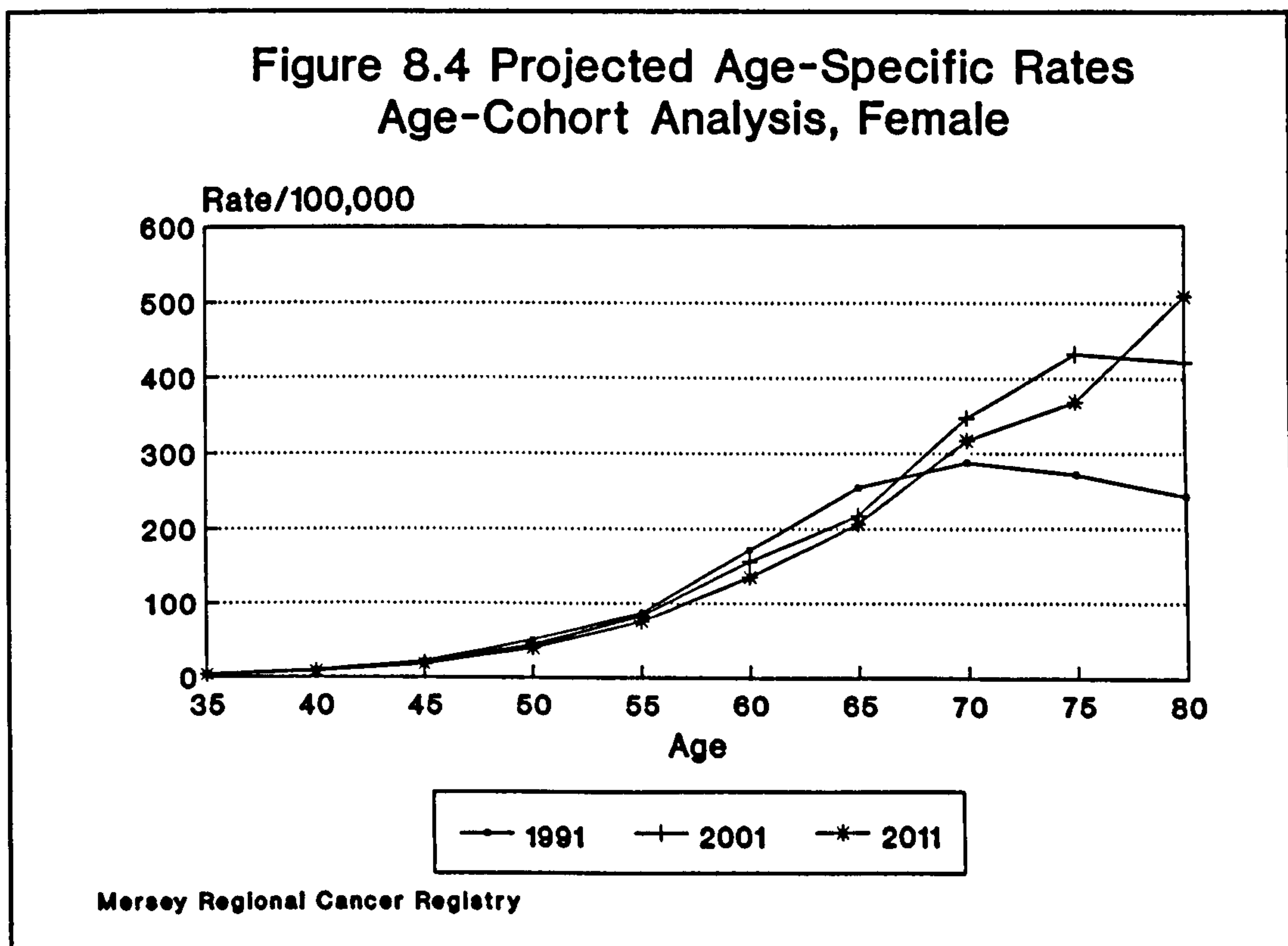
Thus the projected age-specific rate for the 34-39 year age group in 1991 is 4.1 per 100,000.



The age-specific rates for males are projected to decrease over the next twenty years for all age groups (Figure 8.3, Appendix VIII(1)). This is the continuation

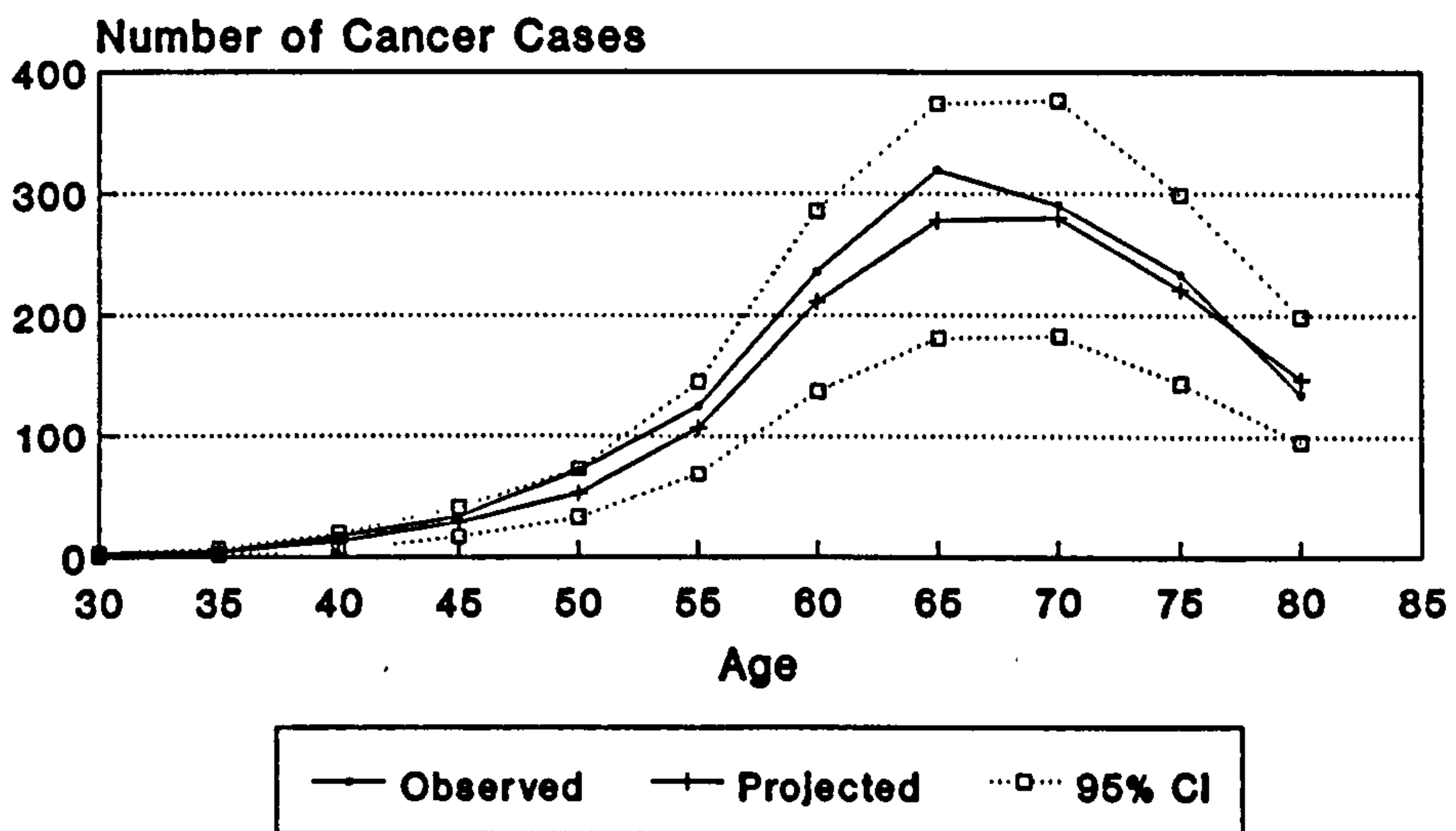


of the trend that started in 1981. The age-specific rates are decreasing over the same time period for females under 70 years of age (Figure 8.4). For females between the ages of 70 years and 79 years, the age specific rates are projected to peak for the quinquennia centred on 2001. For the 80-84 year age group the rates are projected to increase continuously over the time period. For the two oldest groups the age-specific rates are projected to double between 1991 and 2001. This pattern is similar to that observed for males for the quinquennia centred on 1971 through 1981 (Figure 4.6). The confidence intervals for these age specific rates are given in Appendix VIII(2).



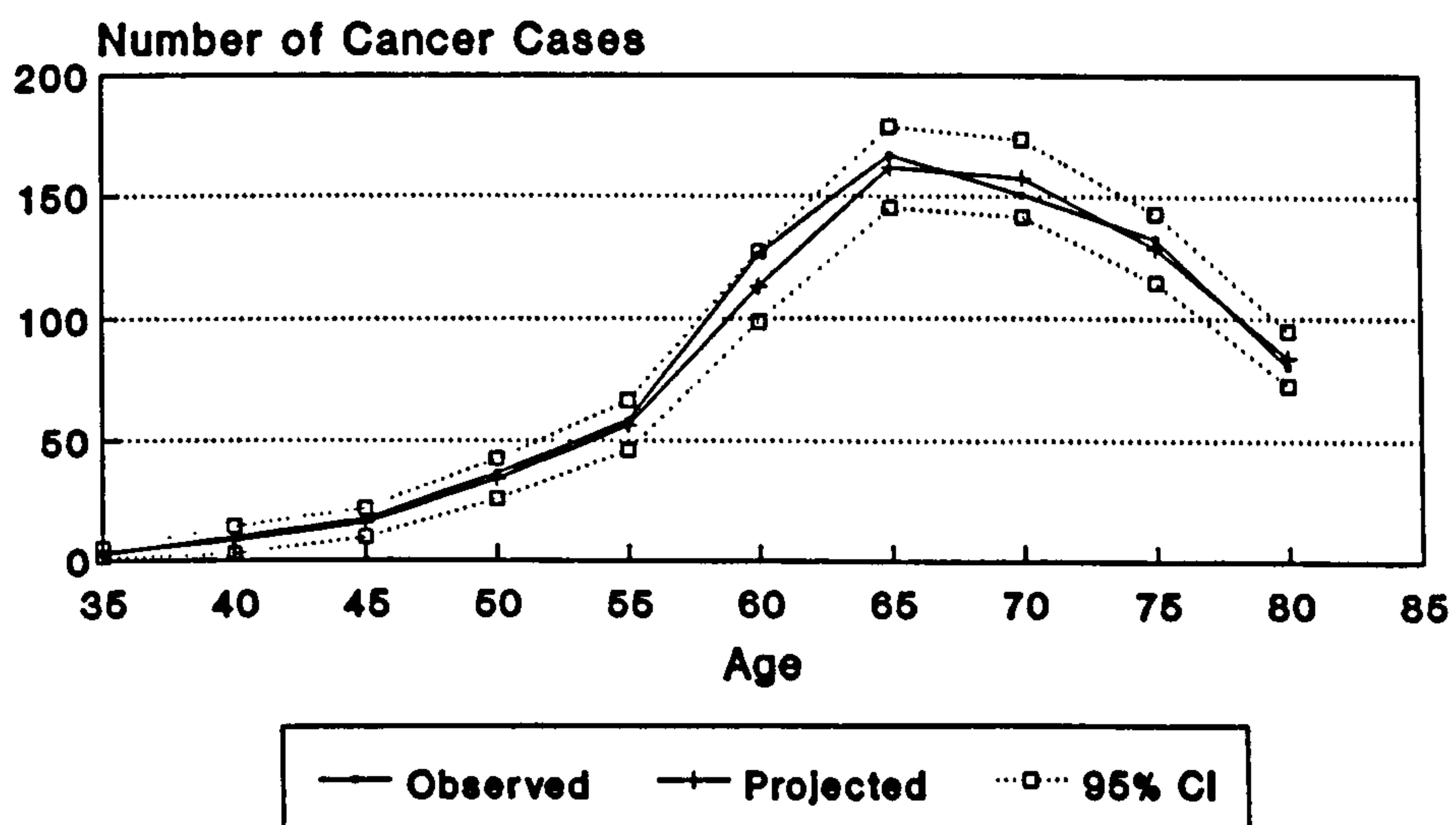
The projected age-specific numbers of cancer cases for the 1991 quinquennia showed good agreement to the

**Figure 8.5 Comparison of Observed and Projected Annual Numbers of Cancer Cases 1991 Quinquennia, Male**



Mersey Regional Cancer Registry  
Based on Age-Period-Cohort Analysis

**Figure 8.6 Comparison of Observed and Projected Annual Numbers of Cancer Case 1991 Quinquennia, Female**



Mersey Regional Cancer Registry  
Based on Age-Cohort Analysis

observed number for both sexes (Figures 8.5 and 8.6). For both sexes, the projected numbers are consistently slightly lower than those observed, but well within the 95% confidence limits for the projections.

### **8.3.3 Health of the Nation Targets**

It is projected that the Health of the Nation target to reduce males lung cancer incidence by 30% between 1991 and 2011 will be met. The projected reduction for age-standardised rates, for the 30-74 year age group, in males is 53%, from 161.3 cases per 100,000 in 1991 to 75.4 cases per 100,000 in 2011. The projected percentage reduction for female lung cancer incidence, for the same age group and time period, is 10%, from 99.3 cases per 100,000 to 88.9 cases per 100,000. This does not meet the Health of the Nation target of 15%, but does reflect an improvement in the burden of lung cancer among this age group. This reduction is a reflection of the decrease in age-specific incidence rates among those women under the age of 70 years.

### **8.3.4 Service Needs**

The projected service needs for males indicate that approximately 100 cases will be treated with either surgery and/or chemotherapy, and almost 200 cases will receive radiotherapy (Table 8.10). This is a projected



Table 8.10 Projected Service Needs for the Year 2001

Sex	Age	Surgery	Chemotherapy	Radiotherapy
Male	30-54	15	19	23
	55-64	33	25	49
	64-74	42	32	86
	75-84	6	14	29
<b>Total</b>		96	90	187
Female	30-54	14	14	17
	55-64	21	25	37
	64-74	27	29	62
	75-84	3	12	22
<b>Total</b>		65	80	134
<b>Total Males and Females</b>		161	170	321

decrease in service needs of over 30% when compared to the figures for 1983-87 (Appendix III(3)). By 2001, 65 females are projected to require surgery, 80 females to require chemotherapy and 134 females to undergo radiotherapy (Table 8.10). This is a slight increase on the service needs for females in 1983-87. The decrease in service needs for males reflects the overall projected decrease in lung cancer incidence for males. The relative stability of the service needs for females despite increasing numbers of cancer cases, reflects the increasing rates in those over the age of 70 years and decreasing rates in those aged less than 70 years. Treatment patterns in 1983-87, on which these projections are based, indicate that higher percentages of younger women receive treatment for lung cancer than the elderly. Thus the increasing numbers of cancer cases in those aged

over 65 years would have little effect on future service needs. Overall, for both males and females, there is projected to be a reduction in service needs between 1983-87 and 1999-2003.

### 8.3.5 Summary of Projections

The results of the model fitting are discussed in Section 8.2.3. These results have shown that the age-specific rates for both male and female lung cancer are influenced by birth cohort. In both cases the incidence is decreasing for the more recent cohorts. However, for the females there will be an absolute rise in the number of lung cancer cases, because of the high rates in the older cohorts. For the males if past trends remain stable, a continuing decline can be expected.

The results also help to explain the lack of fit for the age-standardised rates and the broad-band age-specific rates. In both of these analyses, the trends in smoking behaviours are fitted to changes over calendar period. For the males the trends in tar content of cigarettes would be expected to be a period effect, but quantity smoked, and possibly percentage of the population smoking would be expected to be cohort effects. In the analyses of age-standardised rates and broad-band age-specific rates, the smoking factors that explained the trends the best are the number of cigarettes smoked per person or

per smoker and the percentage of the population smoking for both males and females, together with tar content of cigarettes for males. This is in agreement with the results of the age-period-cohort analyses, which showed that for males both period and cohort variation, while for females there is no period variation, and that the trends could be explained by cohort variation alone.

The projected age-specific incidence for males and females for 1991 agree closely with the observed incidence. For males there is a consistent decrease in the age-specific incidence between 1991 and 2011. The decrease, as expressed by the age-standardised rate for males in the 30-74 year age group, between 1991 and 2011 (53%) is large enough that the Health of the Nation target for the reduction in lung cancer incidence (30%) will be met. For females, there is decreasing lung cancer incidence for those aged under 70 years but a dramatic increase in those aged over 70 years. However, the Health of the Nation target for the reduction of lung cancer incidence by 15% between 1991 and 2011 is for those aged under 75 years. For this restricted age group a 10% reduction in female lung cancer incidence is projected.

Service need are also projected to decrease for males by at least 30% between the quinquennia centred on 1985 and 2001. This reflects the overall decrease in lung cancer incidence. The service needs for females remain fairly



stable over the same period. The large increase in lung cancer incidence among those aged over 65 years is not reflected in increase in service needs because only a small percentage of this age group receives treatment.

Chapter 9

**MODELS BASED ON AGE-PERIOD-COHORT**

**ANALYSIS INCORPORATING INFORMATION ON**

**SMOKING BEHAVIOUR**

The previous chapter demonstrates that the trends in lung cancer incidence for both males and females for Mersey Region are strongly influenced by changes in birth cohort experiences. The trends for males are also affected by changes over calendar period. This chapter will examine the effects of incorporating information on smoking behaviour into the projection models. In particular this chapter will:

- 
- investigate the effect of including smoking variables in the age-period-cohort model.
  - determine the appropriate lag period between changes in the average tar content of cigarettes and changes in lung cancer incidence,
  - determine the best choice of smoking variable to include in the final projection models.
  - describe the scenarios related to the future trends in average tar content of cigarettes and smoking behaviour
  - estimate future values for the smoking variables
  - calculate projected age-specific lung cancer incidence in 1991, 2001 and 2011, using appropriate models and parameter estimates
  - assess whether the Health of the Nation targets for lung cancer are achievable,
  - determine the future service needs in Mersey Region for lung cancer in 2001
-



## 9.1 Data Used

The same lung cancer incidence, and population data are used as in the basic age-period-cohort analyses. That is five year age-specific frequencies for the quinquennia centred on 1961, 1966, 1971, 1976, 1981 and 1986, for the five year age groups from 30-34 years to 80-84 years for males and 35-39 years to 80-84 years for females.

The smoking variables that are investigated are:

- the number of cigarettes smoked per person
- the percentage of the population who smoke
- the number of cigarettes smoked per smoker
- the sales adjusted average tar content of cigarettes

The smoking variables that relate to the quantity of cigarettes smoked and the percentage of the population who smoke are available for each sex separately. Smoking habits are generally established early (Hammond 1966). Therefore smoking data for the age range 25-34 years are used in these analyses to reflect birth cohort behaviour. Whilst smoking habits may change within a cohort over time, (for example a proportion of smokers may quit), these indicators should reflect the smoking behaviour of that birth cohort.

Given that the lung cancer incidence data are averages

for the quinquennia, the smoking data are also included in the model as averages for quinquennia, rather than annual figures. The averages for the quinquennia centred on 1951 through 1986 would include data from 1949-1953 through 1984-1988. Thus the average smoking data, for the age range 25-34 years, would represent the smoking experiences of the cohorts born in 1914-28 through 1949-63 (Table 9.1).

**Table 9.1 Relation between Periods and Cohorts for Age Specific Smoking Information**

<b>Period</b>	1949-1953	1954-1958	...	1984-1986
<b>Age</b>	25-34	25-34	...	25-34
<b>Cohort</b>	1914-1928	1919-1933	...	1949-1963

For the average tar content of cigarettes, the averages for the quinquennia centred on the years 1936, 1941 etc to 1986, are used. For the years 1934 to 1968 the average tar content of cigarettes is only available as 7 year averages. Therefore annual estimates of tar content are calculated from linear interpolation between central points of the intervals (Section 5.1). These annual estimates are used to calculate the five year averages.

There is no age-specific data on the smoking behaviour of the population before 1949. Therefore, there is no cohort-specific information on the smoking behaviour of those cohorts born before 1914 (Table 9.1). Cohorts prior to this time can be represented by the use of indicator

variables (Section 8.1).

## 9.2 Inclusion of Average Tar Content in the Model

### 9.2.1 Methods

Trends in male lung cancer incidence are dependant on period effects, as well as age and cohort effects. Therefore the average tar content of cigarettes is included in the model for males. As with the previous analyses (Sections 6.1 and 7.1) the log of the average tar content of cigarettes is used. The model used to investigate the lag period between changes in the average tar content and lung cancer incidence is:

$$\begin{aligned} \text{Model 5} \quad \ln(R_{ij}) &= \ln(\mu_{ij}) - \ln(N_{ij}) \\ &= \alpha_i + \gamma_k + \tau \ln(\text{tar}_{j-t}) \end{aligned}$$

where  $t$  is the lag period

Five different lag periods are investigated, viz 25 years, 20 years, 15 years, 10 years and 5 years. The contribution of the average tar content for the different lag period are assessed by the size of the change in deviance from the model including age and cohort effects ( $S_{a,t,c;a,c}$ ). The lag period which gives the largest change in scaled deviance is taken as the best lag period for the model.

The analysis in Section 8.2.2 indicates that the trends for female lung cancer do not depend on period effects. Therefore the average tar content is not included in the model at this point.

### 9.2.2 Results

**Table 9.2 Changes in Scaled Deviance for the Different Lag Times for Average Tar Content**

Lag	Change in Scaled Deviance (df = 1)	Final Scaled Deviance (df = 39)
25 years	0.042	83.658
20 years	5.722	77.978
15 years	18.487	65.203
10 years	24.135	59.565
5 years	23.043	60.657

The average tar content of cigarettes has declined from 1934, the year records are first available (Figure 5.8) When included in the modelling its contribution is significant for the lag times of 20 years, 15 years, 10 years and 5 years (Table 9.2). The largest contribution to the fit is for a lag time of 10 years. Examination of the residuals for this model shows them to be randomly distributed.

The pattern of the estimates of the age and cohort parameters are similar to that obtained by the age-period-cohort analyses (Table 9.3 Figures 9.1a and 9.1b).



Table 9.3 Parameter Estimates for Male Lung Cancer from the Age-Cohort Average Tar Modelling (Model 5a)

Age	Parameter (s.e.) Estimate		Cohort	Parameter (s.e.) Estimate	
30-34	-5.161	(0.382)	1874-83	0.000	
35-39	-4.364	(0.361)	1879-88	0.522	(0.097)
40-44	-3.488	(0.348)	1884-93	0.772	(0.092)
45-49	-2.605	(0.338)	1889-98	0.971	(0.092)
50-54	-1.977	(0.329)	1894-1903	1.134	(0.093)
55-59	-1.400	(0.320)	1899-1908	1.162	(0.095)
60-64	-0.898	(0.312)	1904-13	1.165	(0.098)
65-69	-0.556	(0.305)	1909-18	1.115	(0.102)
70-74	-0.295	(0.297)	1914-23	1.172	(0.107)
75-79	-0.147	(0.290)	1919-28	1.130	(0.112)
80-84	-0.033	(0.282)	1924-33	1.067	(0.119)
<b>Average Tar</b>			1929-38	0.850	(0.129)
	0.387	(0.079)	1934-43	0.684	(0.145)
			1939-48	0.609	(0.171)
			1944-53	0.502	(0.221)
			1949-58	-0.157	(0.452)

The final scaled deviance ( $S_{a,t,c} = 59.56$ ,  $df=39$ ) is somewhat larger than that for the age-period-cohort analysis ( $S_{a,p,c} = 57.87$ ,  $df=36$ ). This is not unexpected since the degrees of freedom of the final scaled deviance for Model 5 (39) are greater than those for Model 4 (36). Also the age-period-cohort model (Model 4) places no constraints on the form of the relationship between the

incidence and the period parameters, while the model including tar (Model 5) assumes a linear relationship between the log of the incidence and the log of the average tar content.

### 9.3 Inclusion of Cohort Specific Smoking Information

#### 9.3.1 Methods

Three cohort specific smoking variables are investigated for inclusion in the projection model. These are the number of cigarettes smoked per male or female, the percentage of the males or females smoking and the number of cigarettes smoked per male or female smoker.

The birth cohort smoking variables are approximated by the age-specific data for the age group 25-34 years, which are available for the quinquennia 1951 through 1986 (Section 9.1). Two problems arise from the available data. The first is that the birth cohorts for the smoking variables are 15 years long (1914-28 through 1949-63), while the birth cohorts for the lung cancer incidence data are 10 years long (1874-83 through 1949-58 for males and 1944-53 for females (Tables 8.8)). It would be equally reasonable to ascribe the cohort smoking data to either of the two corresponding sets of ten year birth cohorts for the lung cancer incidence. For example, the birth cohort for the smoking variables for the

quinquennia 1949-53, for the age group 25-34 years would be 1914-28. This would overlap with the birth cohorts 1914-23 and 1919-28, from the lung cancer incidence data (Table 9.4).

The second problem is that there is no information on smoking behaviour for the cohorts 1874-83 to 1949-58 for males and 1874-83 to 1944-53 for females.

<b>Table 9.4 Correspondence between Cohorts for Lung Cancer Incidence and the Smoking Variables</b>			
	<b>Set 1</b>		<b>Set 2</b>
<b>Male</b>			
Lung Cancer Incidence	1919-28 to 1944-53	1914-23 to	1944-53
Smoking Variables	1914-28 to 1939-53	1914-28 to	1944-58
<b>Female</b>			
Lung Cancer Incidence	1919-28 to 1939-48	1914-23 to	1939-48
Smoking Variables	1914-28 to 1934-48	1914-28 to	1939-53

The two problems are addressed simultaneously. In order to investigate the correspondence between the 15 year smoking birth cohort and the 10 year lung cancer incidence cohorts, two sets of analyses are performed. For Set 1 the smoking information corresponds with the later birth cohort for lung cancer incidence. For Set 2 the smoking information corresponds with the earlier

birth cohort for lung cancer incidence (Table 9.4). For example, smoking behaviour for the 1914-28 birth cohort relates to lung cancer incidence for the 1919-28 birth cohort for Set 1 and for the 1914-23 birth cohort for Set 2.

The design matrix for the cohort effects includes indicator variables ( $C_k$  Section 8.1) for those cohorts for which there is no smoking information and one continuous variable that includes the smoking data for those cohorts for which there is information. Thus the design matrix is of the form:

$$X_c = \{C_2, \dots, C_{K'}, S\}$$

where  $C_k$  is the indicator vector for cohort  $k$  ( $k=2, \dots, K'$ ) for which no smoking information is available<sup>1</sup>

$S' = S_c$  information on the smoking variable for those cohorts where it is available, and  
 0 otherwise

<sup>1</sup> This design matrix constrains the cohort parameter estimate for cohort 1 (1874-89) to be equal to 0.



The model is of the form:

$$\begin{aligned} \text{Model 6 } \ln(R_{ij}) &= \ln(\mu_{ij}) - \ln(N_{ij}) \\ &= \alpha_i + \gamma_k + \sigma \ln(sm_k) + \tau \ln(tar_{j-t}) \end{aligned}$$

where  $\gamma_k$  represents the cohort effects for the cohorts for which there is no information on smoking behaviour

and  $\sigma \ln(sm_k)$  represents the cohort smoking effects for the cohorts for which there is information on smoking behaviour

The fit of the models is assessed by the standard procedures outlined in Section 7.1.

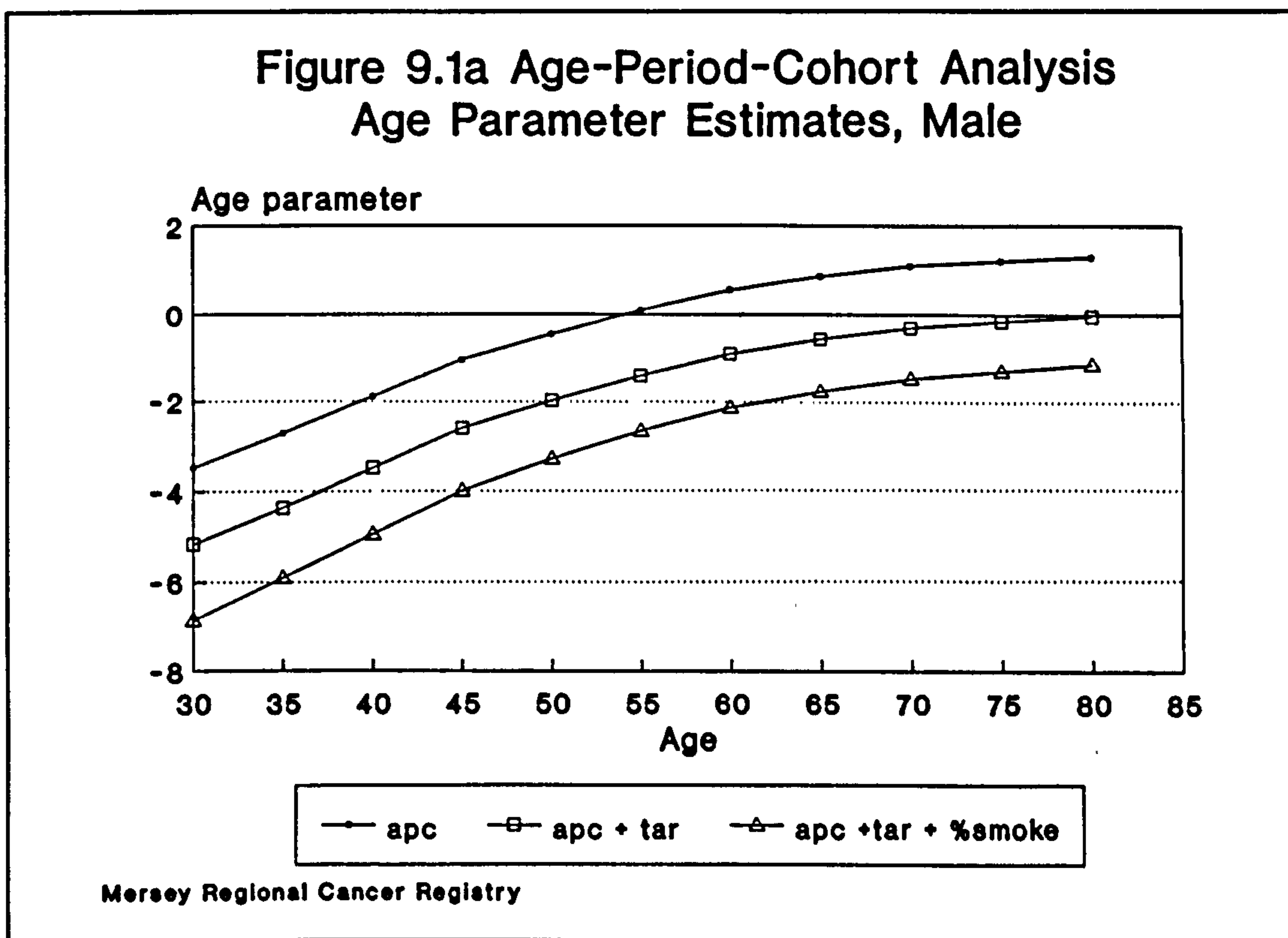
### 9.3.2 Results

Table 9.5 Residual Scaled Deviances for the Models relating Smoking Variables to Male Lung Cancer Incidence

Smoking Variable	Set 1 (df=46) (1914-28 Smoking 1914-23 Cancer)	Set 2 (df=45) (1914-28 Smoking 1919-28 Cancer)
No. Cigs/ Male	130.94	131.73
% who Smoke	118.63	112.75
No. Cigs/ Male Smoker	154.44	150.30

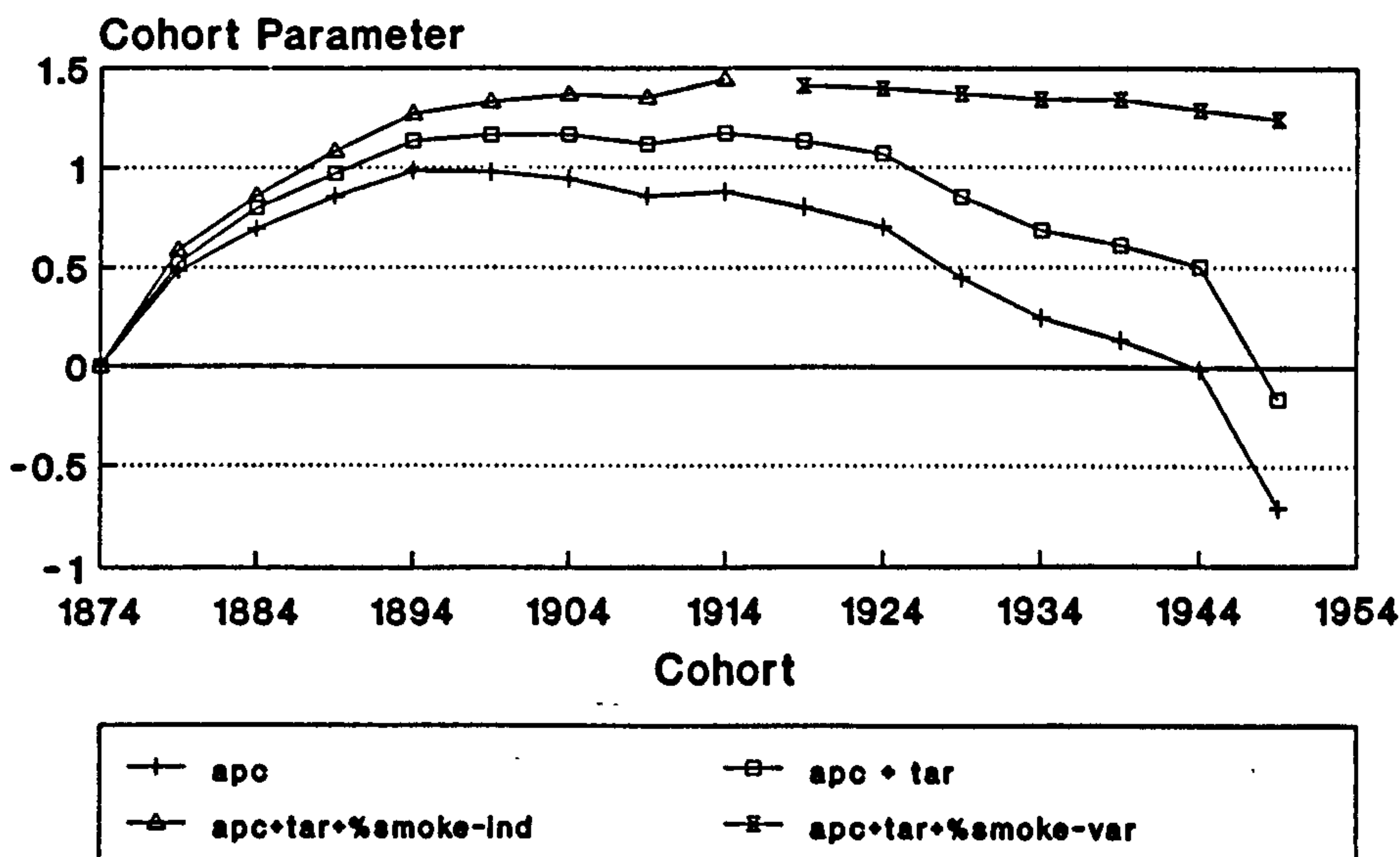
The age indicator variables, the average tar content and the cohort variables for smoking all contribute significantly to the model. The best fit is obtained using the indicator age variables, average tar content of cigarettes lagged by 10 years, and the percentage of

males smoking using the correspondence for Set 2 (i.e. 1914-28 smoking data equivalent to the 1919-28 cancer incidence data) (Table 9.5). This model is referred to as Model 6. The final scaled deviance for Model 6 ( $S_{a,t,ps} = 112.75$ ,  $df=45$ ) is significant. It is also larger than that obtained for Model 5, that including the average tar content of cigarettes and using indicator variables for the cohort effects ( $S_{a,t,c} = 59.56$ ,  $df=39$ ). However, Model 6 allows investigation of changes in smoking behaviour.



The parameter estimates for age for Model 6 show a similar shape to those obtained for Model 5 (Figure 9.1a, Tables 9.3 and 9.6). Also for those cohorts where indicator variables are used the shapes of the parameter estimate curves are similar for the two models (Figure 9.1b, Tables 9.3 and 9.6). Of note is the increase in the

Figure 9.1b Age-Period-Cohort Analyses  
Cohort Parameter Estimates, Male



Mersey Regional Cancer Registry

parameter estimate for the average tar content, which almost doubles with the inclusion of the smoking information. This may be due to the divergence of the cohort parameter estimates for those cohorts where smoking information is available (Figure 9.1b). The variables included in this model do not correspond with those included in any model that fits the age-standardised rates or the broad-band age-specific rates. The model including the number of cigarettes smoked per male gave similar parameter estimates for the age, cohort and tar effects (Table 9.7). The coefficient for the number of cigarettes smoked per male is 0.315. This is similar to the value of the coefficient for the number of cigarettes smoked per male (0.363) in the analysis of the age-standardised rates that included both the average tar content of cigarettes and the number of cigarettes smoked

Table 9.6 Parameter Estimates for Male Lung Cancer from the Age, Percentage of Males who Smoke and Average Tar Modelling (Model 6a)

Age	Parameter (s.e.) Estimate		Cohort	Parameter (s.e.) Estimate	
	30-34	-6.853		(0.293)	1874-83
35-39	-5.908	(0.283)	1879-88	0.586	(0.098)
40-44	-4.947	(0.277)	1884-93	0.859	(0.093)
45-49	-3.986	(0.273)	1889-98	1.081	(0.092)
50-54	-3.292	(0.268)	1894-1903	1.271	(0.092)
55-59	-2.663	(0.263)	1899-1908	1.329	(0.093)
60-64	-2.130	(0.257)	1904-13	1.366	(0.095)
65-69	-1.760	(0.250)	1909-18	1.351	(0.098)
70-74	-1.467	(0.244)	1914-23	1.445	(0.100)
75-79	-1.287	(0.238)			
80-84	-1.141	(0.2312)			
<b>Average Tar</b>			<b>Percent Smoke</b>		
	0.700	(0.065)		0.333	(0.025)

per male (Table 6.3). This similarity enhances our confidence in the reliability of the model presented in Table 9.7.



**Table 9.7 Parameter Estimates for Male Lung Cancer from the Age, Number of Cigarettes Smoked per Male and Average Tar Modelling (Model 6b)**

Age	Parameter (s.e.) Estimate		Cohort	Parameter (s.e.) Estimate	
	30-34	-7.006		(0.310)	1874-83
35-39	-6.043	(0.298)	1879-88	0.538	(0.096)
40-44	-5.066	(0.292)	1884-93	0.812	(0.092)
45-49	-4.090	(0.287)	1889-98	1.037	(0.091)
50-54	-3.386	(0.282)	1894-1903	1.230	(0.092)
55-59	-2.750	(0.276)	1899-1908	1.291	(0.093)
60-64	-2.211	(0.269)	1904-13	1.331	(0.096)
65-69	-1.841	(0.262)	1909-18	1.320	(0.099)
70-74	-1.544	(0.256)	1914-23	1.420	(0.102)
75-79	-1.361	(0.250)			
80-84	-1.215	(0.243)			
<b>Average Tar</b>			<b>No. Cigs/ Male</b>		
	0.736	(0.068)		0.315	(0.025)

For females the best fit for the smoking data is with Set 2 (i.e 1914-28 smoking data corresponding to the 1919-28 cancer incidence data) and the use of the percentage of females smoking (Table 9.8). The final scaled deviance ( $S_{a,ps} = 52.02$ ,  $df=41$ ) was marginally significant.

Inclusion of indicator variables for period effects contributed significantly to the model ( $S_{a,p,ps;a,ps} = 11.24$ ,

df = 5). This led to the investigation of the inclusion of the average tar content into the model lagged by either 10 years, 15 years or 20 years. As in the investigation for the males the lag period is evaluated by the relative change in the scaled deviance.

**Table 9.8 Residual Scaled Deviances for the Models relating Smoking Variables to Female Lung Cancer Incidence**

Smoking Variable	Set 1 (df=42) (1914-28 Smoking 1914-23 Cancer)	Set 2 (df=41) (1914-28 Smoking 1919-28 Cancer)
% who Smoke	81.92	52.02
No. Cigs/Female	71.76	76.10
No. Cigs/Female Smoker	73.71	82.20

If the average tar content is included in the model with the percentage of females who smoke the best fit is obtained with a lag period of 10 years. This lag period is the same as that for males. The final scaled deviance ( $S_{a,t,ps} = 42.43$ ,  $df = 40$ ), indicates an extremely good fit. The parameter estimate for average tar content for females is 0.3587, which is approximately half that for males (Table 9.9). Females have tended to smoke lower tar cigarettes than males, and therefore changes in tar content would have a smaller effect on females than males. The parameter estimate for the percentage of females who smoke is 0.582, which is almost double that for males. This may be a reflection of the distribution of smoking patterns among females compared to males. When

the age parameter estimates and the cohort parameter estimates are compared with those obtained from the age-period-cohort analysis there is good agreement (Figures 9.2a and 9.2b).

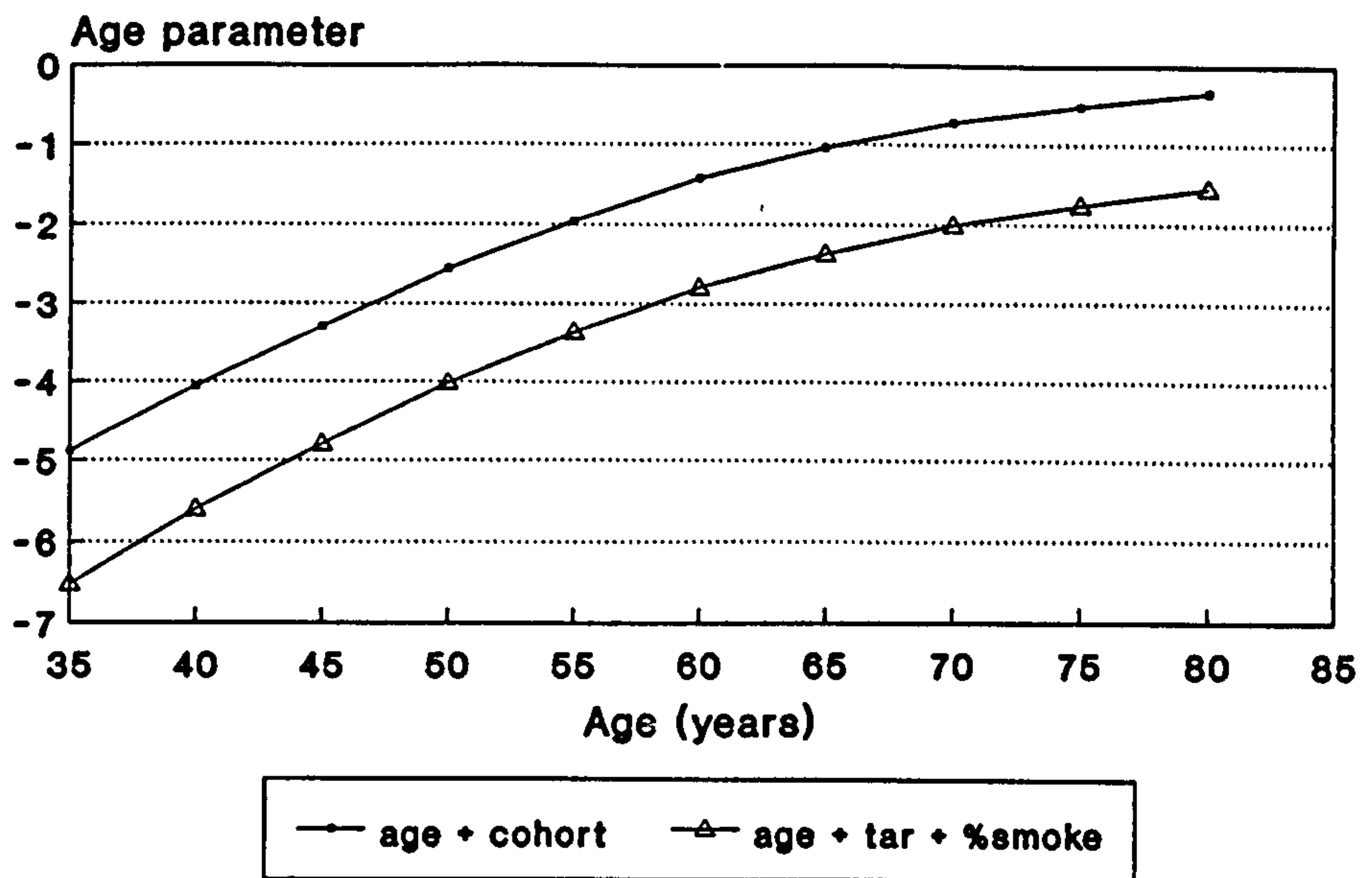
**Table 9.9 Parameter Estimates for Female Lung Cancer from the Age, Percent Smoking and Average Tar Modelling (Model 6a)**

Age	Parameter (s.e.) Estimate		Cohort	Parameter (s.e.) Estimate	
35-39	-6.523	(0.512)	1874-83	0.000	
40-44	-5.604	(0.504)	1879-88	0.239	(0.150)
44-49	-4.799	(0.497)	1884-93	0.284	(0.144)
50-54	-4.024	(0.489)	1889-98	0.522	(0.140)
55-59	-3.371	(0.479)	1894-1903	0.836	(0.140)
60-64	-2.802	(0.466)	1899-1908	1.077	(0.143)
64-69	-2.377	(0.451)	1904-13	1.411	(0.149)
70-74	-2.010	(0.439)	1909-18	1.743	(0.155)
74-79	-1.766	(0.427)	1914-23	2.048	(0.163)
80-84	-1.553	(0.414)			
<b>Average Tar</b>			<b>Percent Smoke</b>		
	0.359	(0.116)		0.582	(0.045)

For the model including the number of cigarettes smoked per female (Table 9.10), the parameter estimate for average tar content (0.6389) is approximately equivalent to that for males (0.7361), while the coefficient for the number of cigarettes smoked per female (0.6935) approximately double that of the equivalent coefficient

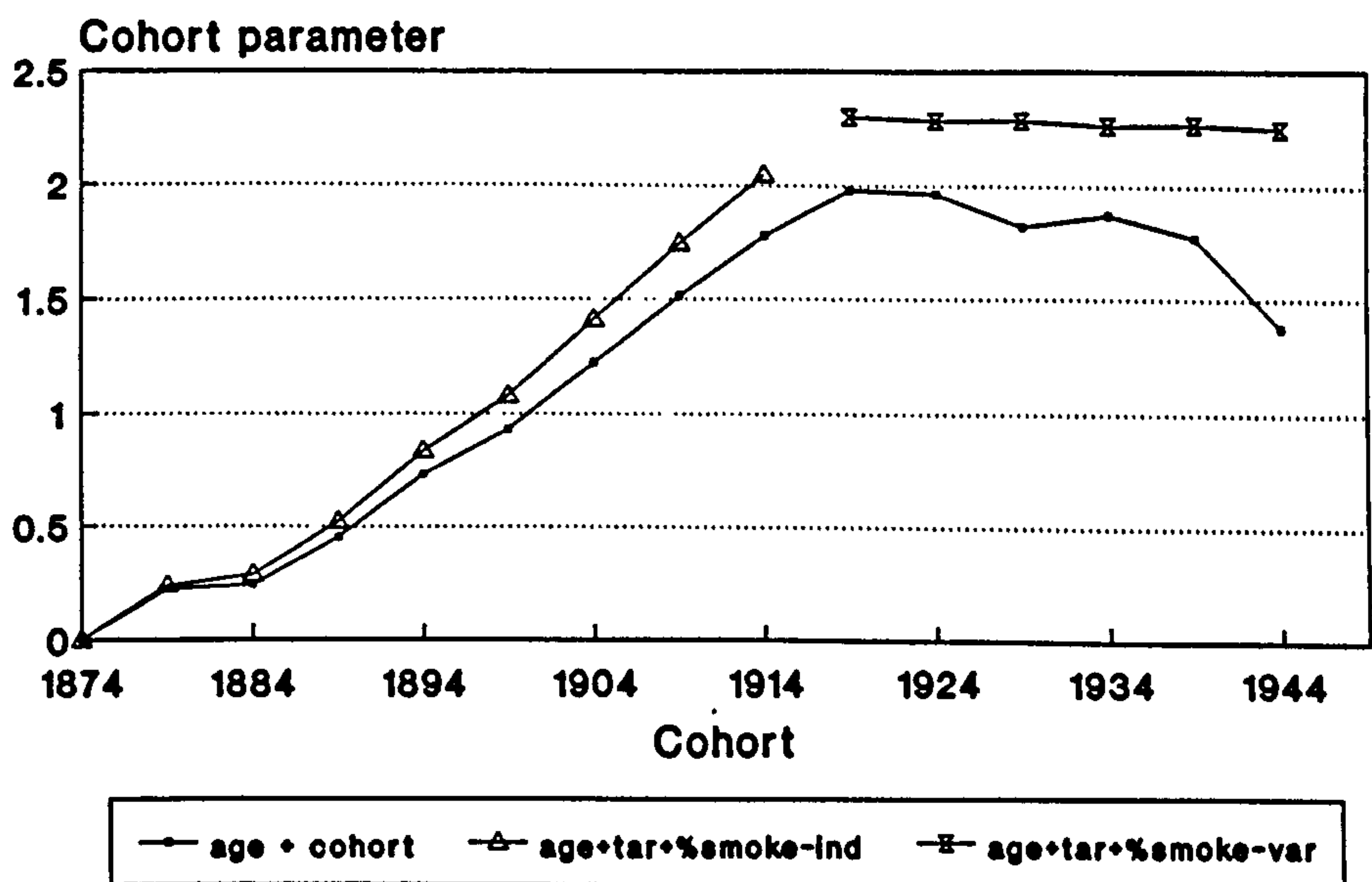


Figure 9.2a Age-Period-Cohort Analyses  
Age Parameter Estimates, Female



Mersey Regional Cancer Registry

Figure 9.2b Age-Period-Cohort Analyses  
Cohort Parameter Estimates, Female



Mersey Regional Cancer Registry

for males (0.3153). This again may be a reflection of the difference in the distribution of smoking habits among males and females. As expected the shape of the age and



Table 9.10 Parameter Estimates for Female Lung Cancer from the Age, Number of Cigarettes Smoked per Female and Average Tar Modelling (Model 6b)

Age	Parameter (s.e.) Estimate		Cohort	Parameter (s.e.) Estimate	
35-39	-7.784	(0.608)	1874-83	0.000	
40-44	-6.780	(0.596)	1879-88	0.144	(0.143)
44-49	-5.903	(0.585)	1884-93	0.207	(0.137)
50-54	-5.072	(0.573)	1889-98	0.466	(0.135)
55-59	-4.369	(0.559)	1894-1903	0.804	(0.137)
60-64	-3.749	(0.543)	1899-1908	1.079	(0.144)
64-69	-3.306	(0.527)	1904-13	1.441	(0.153)
70-74	-2.906	(0.512)	1909-18	1.808	(0.162)
74-79	-2.633	(0.497)	1914-23	2.149	(0.173)
80-84	-2.400	(0.482)			
<b>Average Tar</b>			<b>No. Cigs/Female</b>		
	0.639	(0.134)		0.694	(0.055)

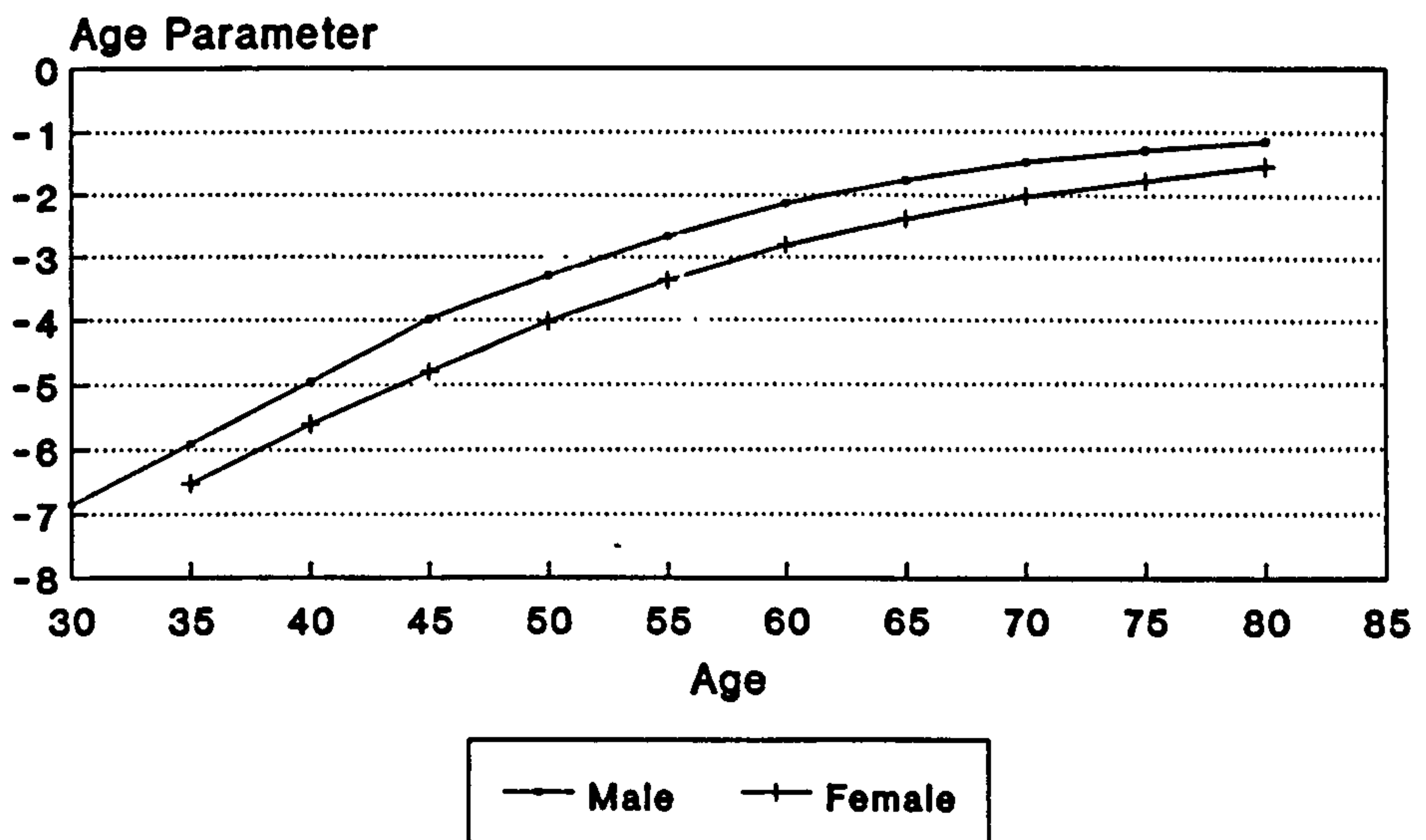
cohort parameter curves are similar for the two models

The inclusion of both the percentage of the population who smoke and the number of cigarettes smoked per person in the model is investigated. For males, both smoking variables contribute significantly to the model. However, the model including both variables would indicate decreasing lung cancer incidence with increasing number of cigarettes smoked per male. This is contrary to known evidence (Doll and Peto 1976). Therefore this model is

not considered further. For females if both the percentage of females who smoke and the number of cigarettes smoked per female are included in the model, the number of cigarettes smoked per female does not contribute significantly to the model ( $S_{a,t,ps,np;a,t,ps} = 0.01$ ,  $df = 1$ ) and thus this model can be discarded for projections.

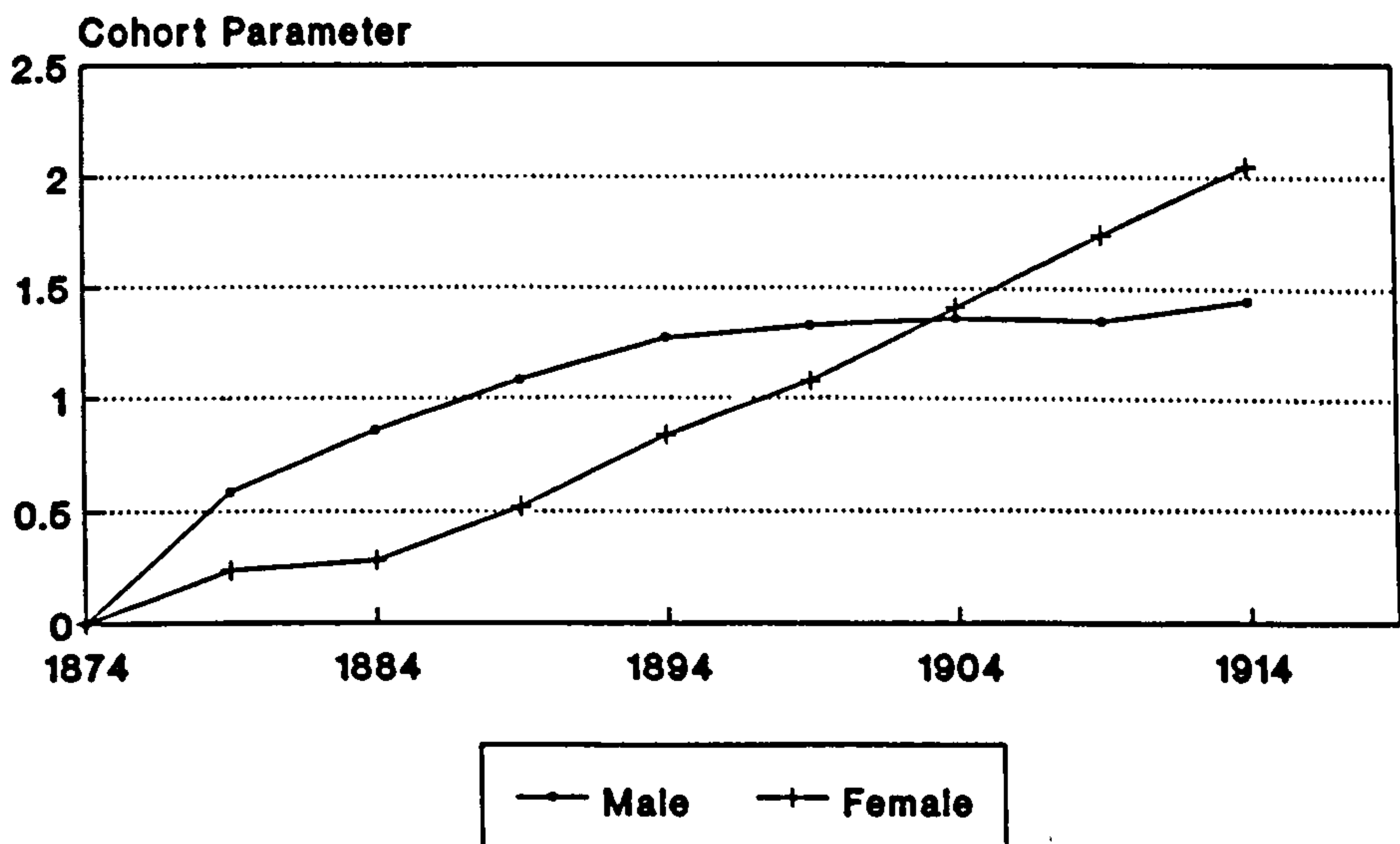
Examination of the age and cohort parameter estimate plots show that the age parameters for males and females are roughly parallel. The plot of the cohort parameter estimates indicate that the early female cohort parameter estimates increase more slowly than those for males. However, the parameter estimates for male cohorts born after the turn of the century have plateaued, while those for females continue to increase (Figures 9.3a and 9.3b).

**Figure 9.3a Age Parameter Estimates  
for Model 6 including % who Smoke,  
Males and Females**



Mersey Regional Cancer Registry

**Figure 9.3b Cohort Parameter Estimates  
for Model 6 including % who Smoke,  
Males and Females**



Mersey Regional Cancer Registry

## 9.4 Projections

The previous analyses have shown that the trends for both male and female lung cancer incidence rates are related to the trends in average tar content of cigarettes and the percentage of the male or female population who smoke, or the average number of cigarettes smoked per male or female. The future age-specific incidence rates can be determined by the equation:

$$\begin{aligned} \text{Model 6 } \ln(R_{ij}) &= \ln(\mu_{ij}) - \ln(N_{ij}) \\ &= \alpha_i + \gamma_k + \sigma \ln(sm_k) + \tau \ln(tar_{j-t}) \end{aligned}$$

where  $\gamma_k$  represents the cohort effects for cohorts from 1874-83 through 1914-23

and  $\sigma \ln(sm_k)$  represents the cohort smoking effects for cohorts after 1914-23

This model is based on the assumption that the age parameters remain constant over the period of projection and that the period and cohort effects are additive.

In order to make the projections for the years 1991, 2001 and 2011, the following estimates are required: the average tar content of cigarettes for the years 1981, 2001 and 2011, and the percentage of the population who smoke and the number of cigarettes smoked per person for the birth cohorts born after 1959. These estimates are discussed in Section 9.4.1 and 9.4.2. The age parameter



estimates and cohort parameter estimates for the cohorts 1874-83 through 1914-23 are determined from the appropriate analysis (Model 6) (Tables 9.6, 9.7, 9.9 and 9.10).

#### 9.4.1 Determination of the Average Tar Content

In the analyses a lag period of 10 years for the average tar content gave the best fit. Therefore, in order to project lung cancer incidence for 1991, 2001 and 2011, it is necessary to determine the average tar content of cigarettes for 1981, 1991 and 2001. The average tar content is available for 1981, but due to the cessation of the publication of the market share of cigarette brands, those for 1991 and 2001 have to be determined. Two sets of estimates are used:

- |   |
|---|
| <p>t1&gt; estimates based on extrapolation of the log linear trend in tar content to 1991 and 2001,<br/>t2&gt; estimates based on assuming that the tar level remains constant after 1987</p> |
|---|

These two sets of estimates are chosen to reflect a continuation of the decline in the average tar content of cigarettes, which could be considered to be the best case scenario, and no future decline which could be considered to be the worst case scenario.

#### 9.4.2 Determination of the Information on Smoking Behaviour

Information on smoking behaviour needs to be estimated for the more recent cohorts. These are the birth cohorts 1954-63 through 1974-83 for males, and through 1949-58 through 1969-78 for females (Table 8.8). The methods used are:

- c1> to estimate the values of the smoking variables by linear extrapolation of the log of the smoking variables
- c2> to assume that the more recent birth cohorts will achieve the stated goal for the Health of the Nation; viz 20% smoking and a reduction of 40% in the number of cigarettes smoked.

These two scenarios could be considered to be the best case, achievement of the Health targets, and the intermediate case, continuation of the present trends.

#### 9.4.3 Determination of the Age-specific Lung Cancer Incidence for 1991, 2001 and 2011

The future age-specific lung cancer incidence for the years 1991, 2001 and 2011 are determined from the equation for Model 6. Eight scenarios are investigated, for each sex, using the different estimates of future smoking parameters outlined above (Table 9.11). These scenarios will allow investigation of the effects on lung

cancer incidence of the continuation of current trends in the three smoking variables, as well as the effect of attaining the Health of the Nation targets. They may also inform on further measures that are necessary to achieve the Health of the Nation target for lung cancer. Also of interest will be the effect of the different scenarios on service needs for lung cancer.

**Table 9.11 The Scenarios used for Projections**

<b>Scenario</b>	<b>Tar Content</b>	<b>Percent Smoking</b>
9.1	t1 Extrapolation	c1 Extrapolation
9.2	t1 Extrapolation	c2 Health of the Nation
9.3	t2 Constant (1986)	c1 Extrapolation
9.4	t2 Constant (1986)	c2 Health of the Nation
	<b>Tar Content</b>	<b>Cigarettes/Person</b>
9.5	t1 Extrapolation	c1 Extrapolation
9.6	t1 Extrapolation	c2 Health of the Nation
9.7	t2 Constant (1986)	c1 Extrapolation
9.8	t2 Constant (1986)	c2 Health of the Nation

For each scenario the age-specific frequencies and rates are presented for 1991, 2001 and 2011. Confidence intervals are calculated for the age-specific frequencies using the methods proposed by Hakulinen and Dyba (1994). For ease of comparison among the different scenarios, the



age-standardised rates are also presented, standardised to the 1991 Mersey Region population aged between 30 years and 84 years for males, and 35 years and 84 years for females.

In order to determine the fit of the projection equations, the projected age-specific frequencies for 1991 are compared to those actually observed in Mersey Region. The age-standardised rates for the age groups 30-74 for males and 35-74 for females are calculated, for the years 1991 and 2011 using the Mersey region population for 1991 as the standard (Section 8.3.1). The projected percentage change in lung cancer incidence is calculated for each scenario to determine whether the Health of the Nation target for lung cancer will be met. The age-specific frequencies of lung cancer incidence are used to determine service needs for the years 2001. These are calculated using the age-specific treatment pattern for each sex recorded for 1983-87 (Williams et al 1993).



## 9.5 Results

### 9.5.1 Average Tar Content

The average tar content of cigarettes fell steeply after 1969, when annual levels were first available. After 1973 the trend levelled out and examination of the plot of the log of the average tar content showed that this trend appears linear (Figure 5.8). Thus, the projection equation for the annual average tar content is based on the log of the average tar content for the years 1973 to 1987. The resulting equation is:

$$\text{Log(Average Tar)} = 4.893 - 0.0262 (\text{Year}-1900)$$

This is equivalent to a 2.62% decline in tar content per year. The average tar content of cigarettes for the years of interest are given in Table 9.12.

Table 9.12      Estimates of the  
Average Tar Content  
of Cigarettes

Quinquennia	Tar Content
1981 <sup>1</sup>	15.9 mg/cig
1986 <sup>1</sup>	14.1 mg/cig
1991	12.3 mg/cig
2001	9.5 mg/cig

<sup>1</sup> Observed

Given the lag of 10 years between the trends for average tar content and the trends in lung cancer incidence, the average tar content of cigarettes for 1981, 1991 and 2001

would correspond to lung cancer incidence in 1991, 2001 and 2011 respectively.

### 9.5.2 Extrapolation of the Cohort Smoking Values

The percentage of males who smoke declines over the cohorts of interest (Figure 5.4). The regression of the log of the percentage of males smoking for the age group 25 to 34 years of age against cohort gave the equation:

$$\log(\%sm_k) = 4.363 - 0.084 (k-9)$$

where  $k=1$  for the 1874-83 birth cohort etc

The number of cigarettes smoked per male increases for the cohorts 1914-28 to 1924-33 and then plateaus until the cohort for 1934-48. After the 1934-48 cohort the number of cigarettes smoked per male decreases (Figure 5.2). Thus the projection equation for the number of cigarettes smoked per male is based only on the cohorts 1934-48 through 1949-1963.

This gives the equation:

$$\log(np_k) = 4.616 - 0.156 (k-13)$$

The estimates of the smoking variables for males are given in Table 9.13.

The percentage of females who smoke declines over the

Table 9.13

## Estimates of the Smoking Variables for Males

Cohort	Estimate of % Males who Smoke	No. Cigs/Male
1959-68	36.8%	46
1964-73	33.8%	40
1969-78	31.1%	34
1974-83	28.6%	29

cohorts of interest (Figure 5.4), and the number of cigarettes smoked per female increases to the 1939-48 cohort and then begins to decline (Figure 5.2). The regression of the log of the percentage of females who smoke for the age group 25 to 34 years of age against cohort gives the equation:

$$\log(\%sm_k) = 4.035 - 0.0378 (k-9)$$

The regression equation relating the number of cigarette smoked per female for the cohorts from 1939-48 to 1954-63 is:

$$\log(np_c) = 4.187 - 0.0879 (k-13)$$

The estimates of the female smoking variables are given in Table 9.14.

These results indicate that the decline in smoking is less steep among the females than among the males and

Table 9.14 Estimates of the Smoking Variables for Female

Cohort	Estimate	
	% Females who Smoke	No. Cigs/Female
1959-68	40.2%	42
1964-73	38.7%	39
1969-78	37.3%	36
1974-83	35.9%	33

that for the more recent cohorts a higher proportion of females are projected to smoke than their male counterparts.

### 9.5.3 Health of the Nation Targets for Smoking

The Health of the Nation target for the percentage of the population who smoke is:

To reduce the prevalence of cigarette smoking in men and women aged 16 and over to no more than 20% by the year 2000.

The Health of the Nation target for the quantity smoked is:

To reduce the consumption of cigarettes by at least 40% by the year 2000 (from 98 billion manufactured cigarettes per year in 1990 to 59 billion)

This would translate to targets of 32.5 cigarettes per male per week and 26.5 cigarettes per female per week.



The target for the percent of the population who smoke is considerably lower than that projected to occur, if present trends continue. The target for the number of cigarettes smoked per male will almost be met by the year 2001, but will not be met for the number of cigarettes smoked per female.

The projections based on the Health of the Nation targets assume that only those cohorts from 1959-69 through 1974-83 achieve this target. The rationale for this assumption is that the percentage of the population who smoke in the 25-34 year age group is used as an indicator for smoking behaviour. This indicator does not explicitly account for changes in quitting behaviour after the age of 35 years. If the earlier cohorts (before 1959-68) increase their rate of quitting, it is not possible to incorporate this in the present models.

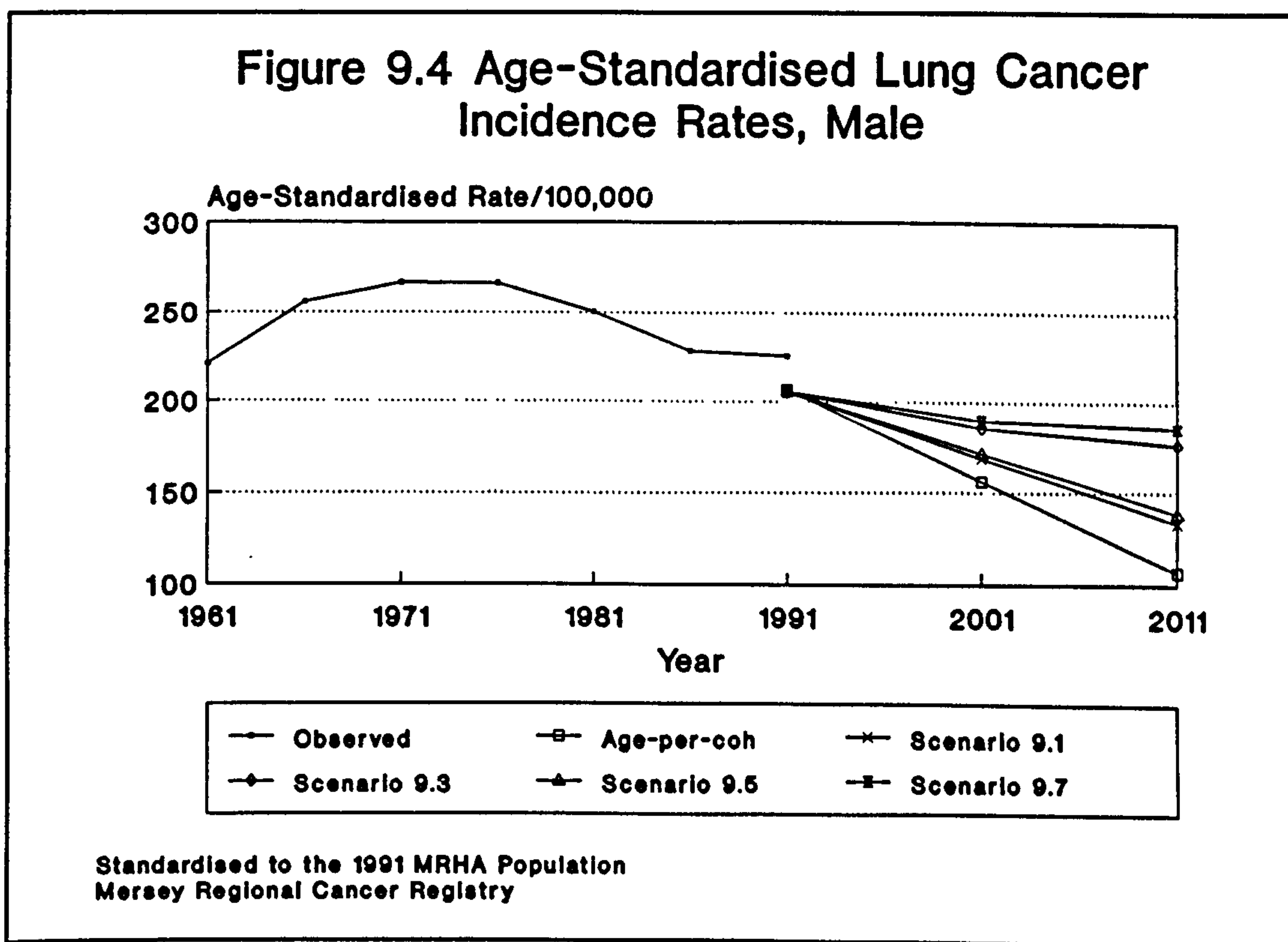
#### **9.5.4 Projected Lung Cancer Incidence**

This section presents the projected age-specific lung cancer incidence, for males and females separately. The effect of the different scenarios on the age-specific frequencies and rates are presented.

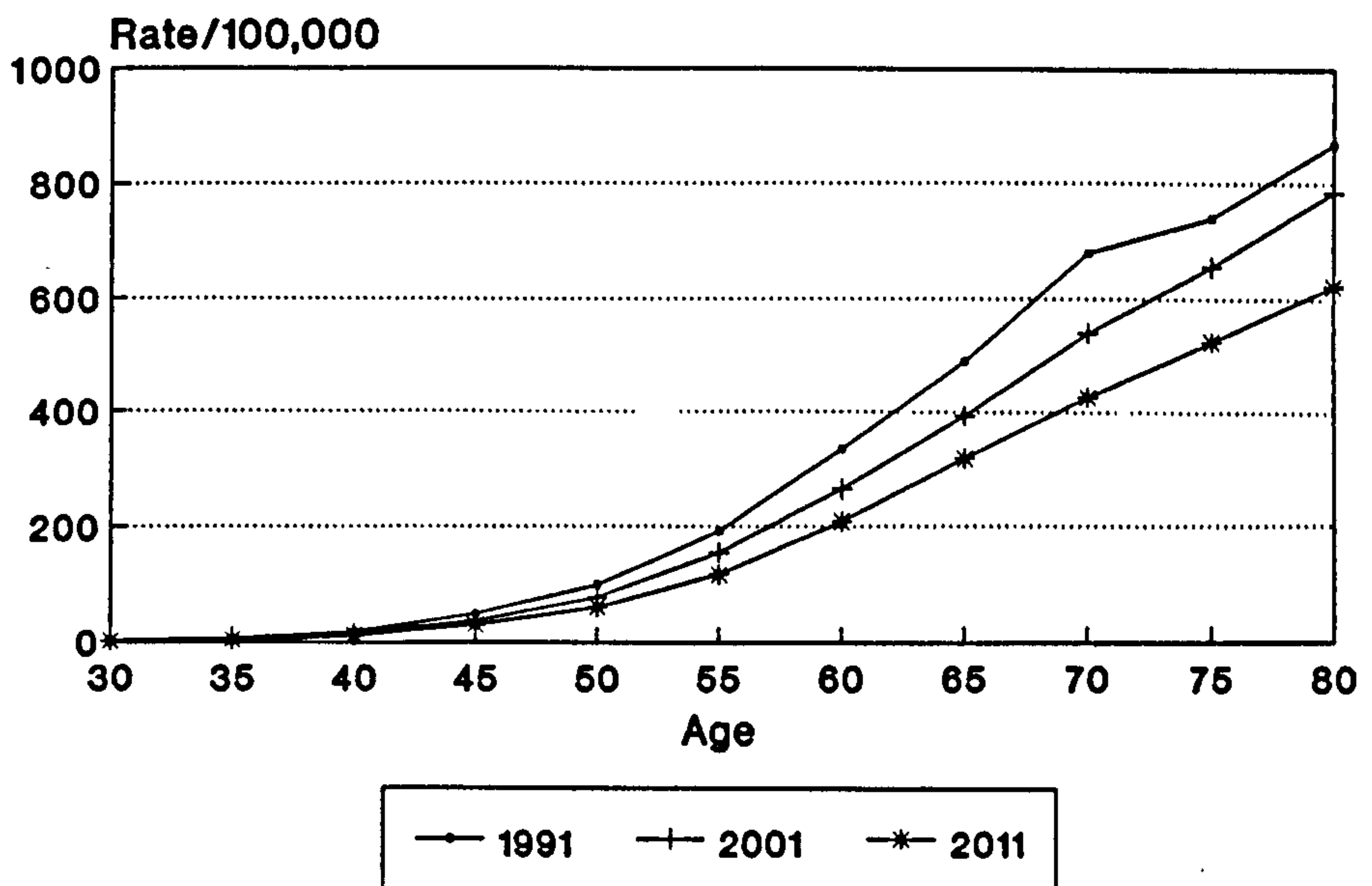
In comparing the eight different scenarios, it should be noted that for 1991 all four scenarios based on the percentage of the population smoking (Scenarios 9.1-9.4) will produce the same estimates for the age-specific frequencies for lung cancer incidence. The same is true for those scenarios based on the average number of cigarettes smoked per person (Scenarios 9.5-9.8). Since these estimates are based on available data for both the average tar content and the smoking variable, they are not subject to variation in the assumptions relating to these variables. When the estimates, for 2001, of the age-specific lung cancer incidence based on the extrapolation of the smoking variables are compared to those estimates based on the Health of the Nation targets the only difference is for the age groups 30-34 years and 35-39 years. Similarly for the projections for 2011, the only differences between these pairs of scenarios will be for the age groups 30-34 years through 45-49 years (Appendices IX(1) and IX(2)).

The age-standardised lung cancer incidence for males is projected to decline for all scenarios (Figure 9.4). These decreases are also apparent for the age-specific lung cancer incidence rates (Figures 9.5a, 9.5b, 9.5c and 9.5d). The decline is less marked for those scenarios in which the average tar content of cigarettes is assumed to remain at the 1986 level. The effect of attaining the Health of the Nation targets for smoking behaviour will

be to reduce the incidence in the younger age groups. In these age groups the incidence rates are small (less than 50/100,000) and thus the overall effect on the total number of projected incident cases for 2001 and 2011 is limited. There is also little difference between the frequencies based on percentage of males who smoke and those based on the average number of cigarettes smoked per male.

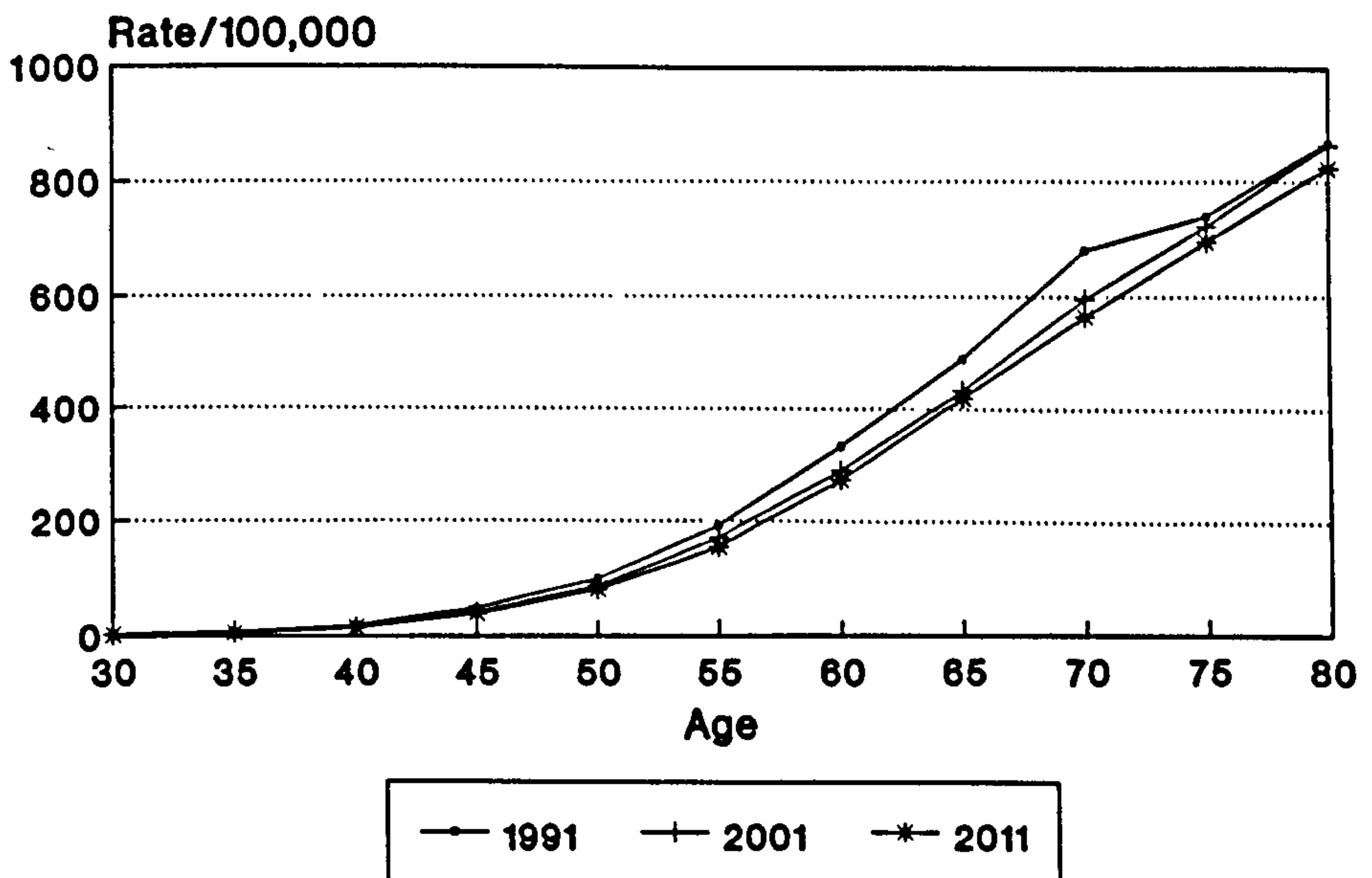


**Figure 9.5a Projected Lung Cancer Incidence Rates Scenario 9.1, Male**



Mersey Regional Cancer Registry

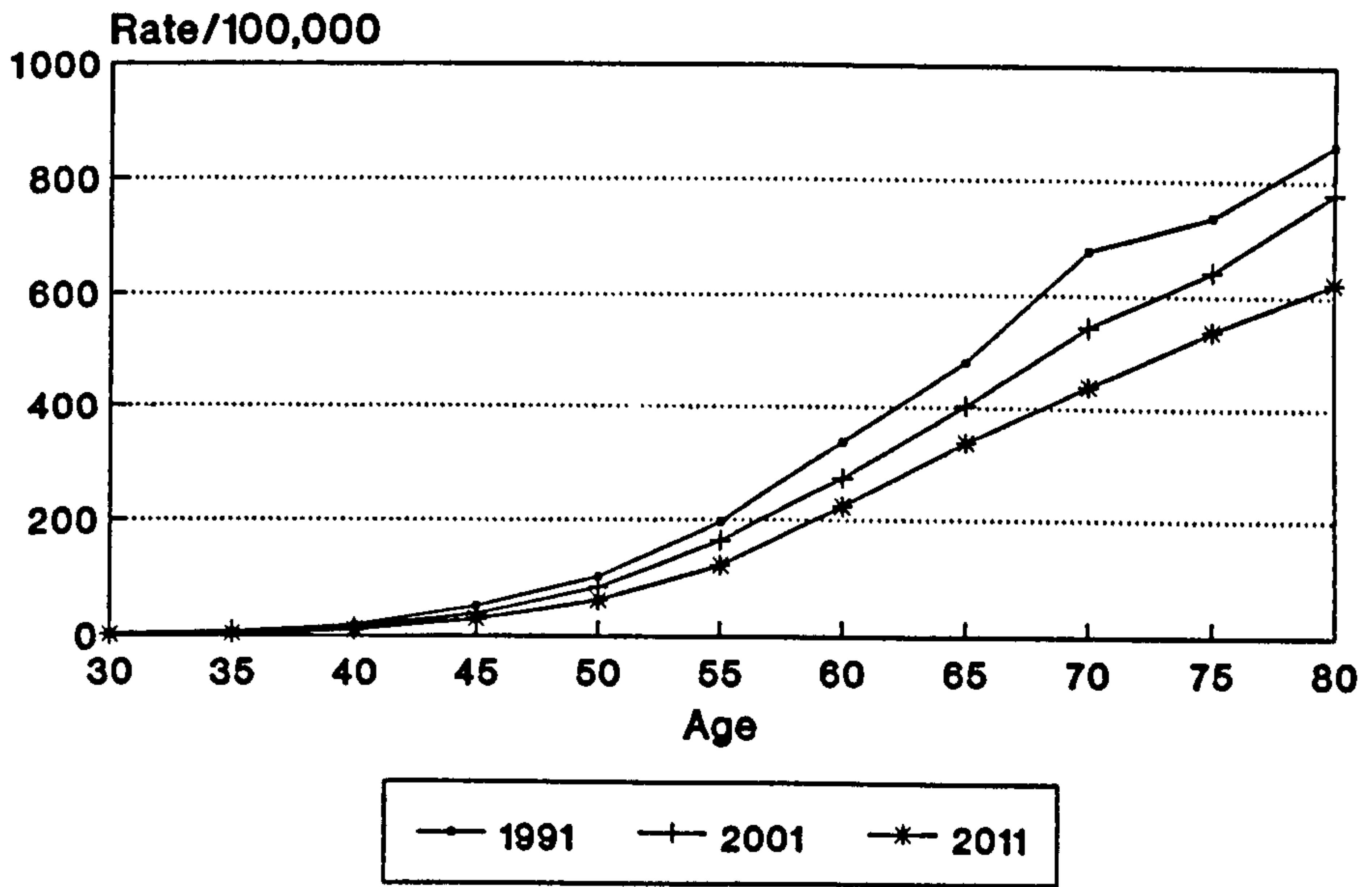
**Figure 9.5b Projected Lung Cancer Incidence Rates Scenario 9.3, Male**



Mersey Regional Cancer Registry

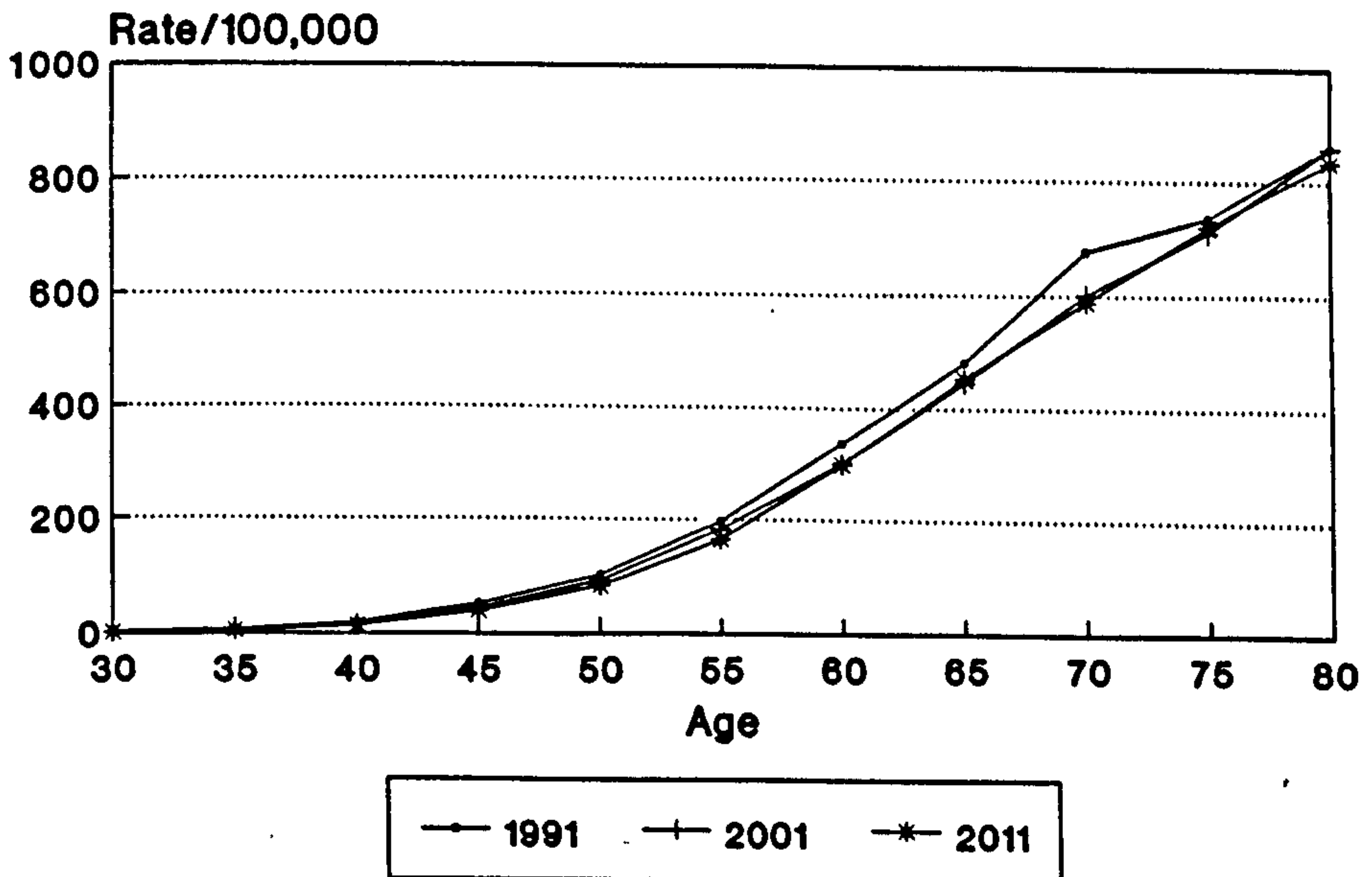


**Figure 9.5c Projected Lung Cancer Incidence Rates Scenario 9.5, Male**



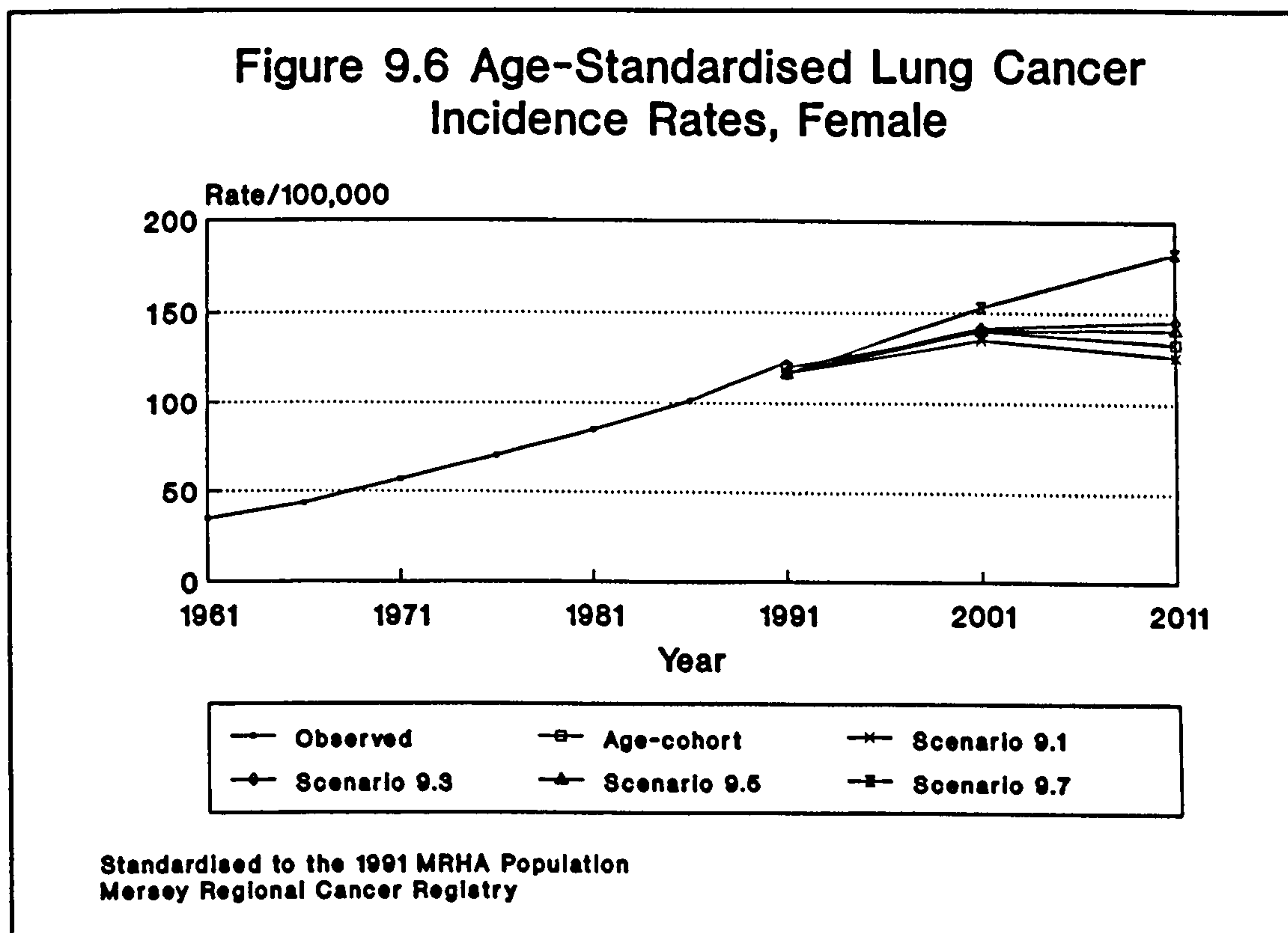
Mersey Regional Cancer Registry

**Figure 9.5d Projected Lung Cancer Incidence Rates Scenario 9.7, Male**

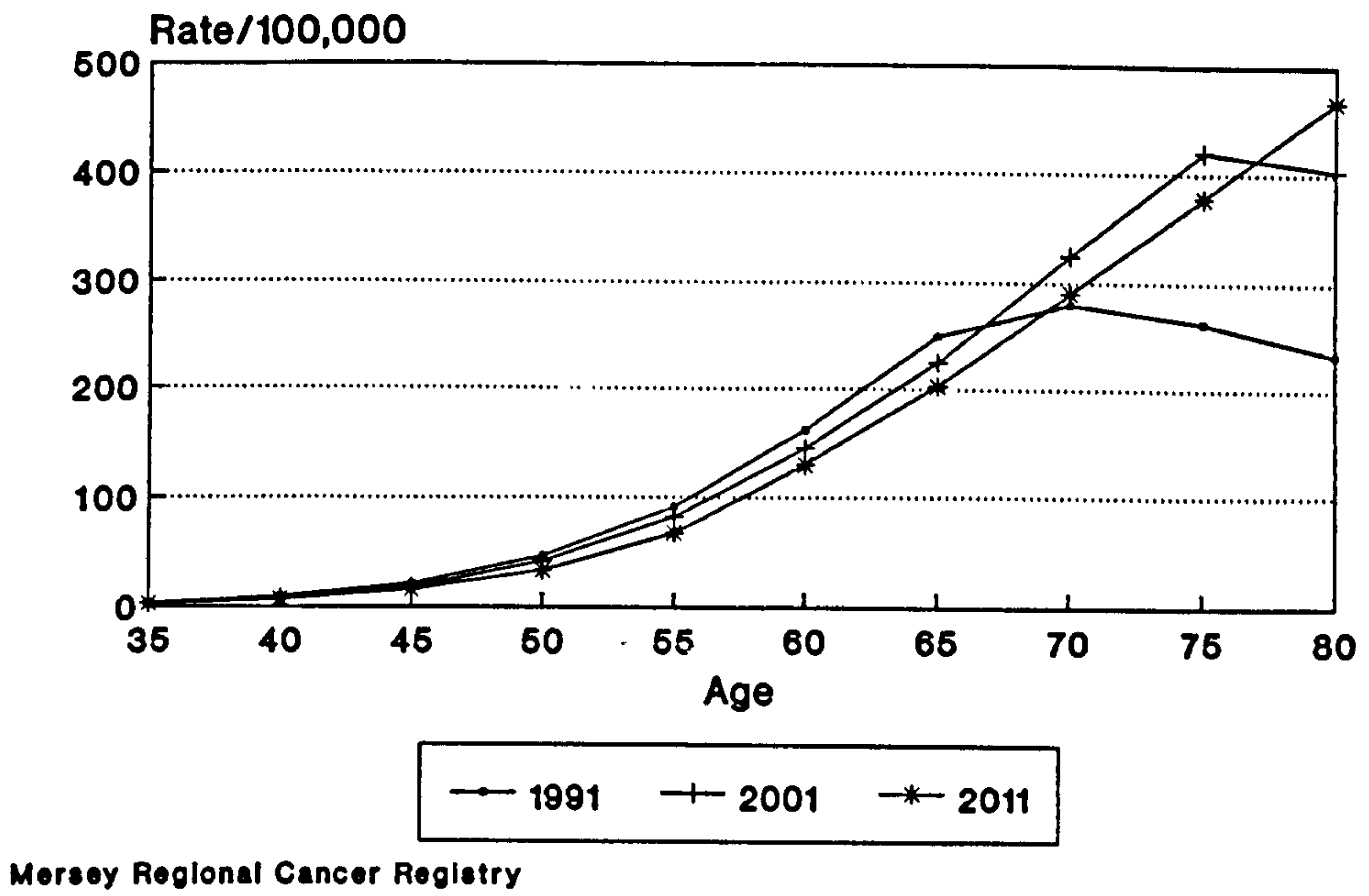


Mersey Regional Cancer Registry

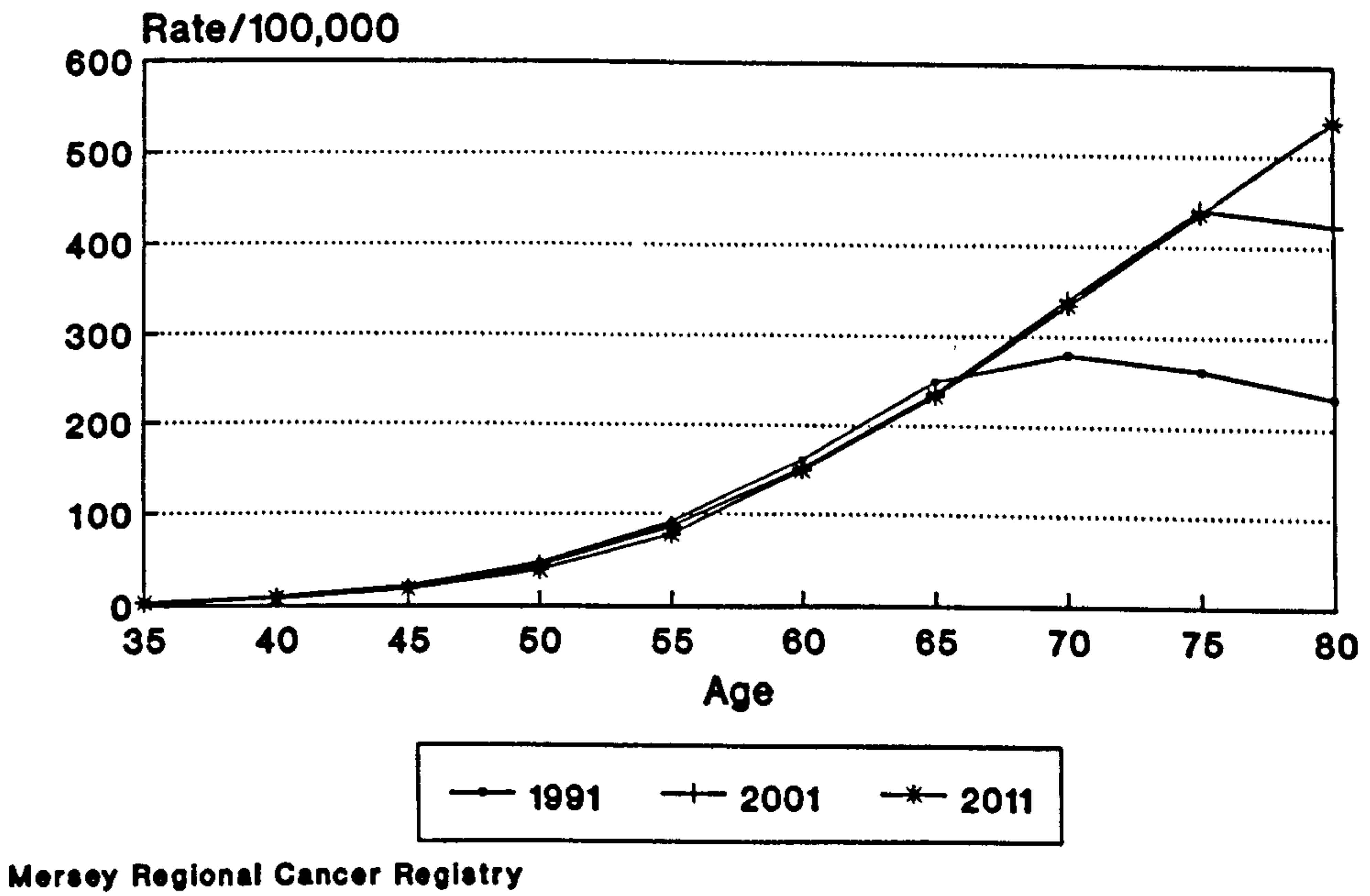
Unlike male lung cancer incidence, the age-standardised female lung cancer incidence is still continuing to increase. The projections based on the extrapolation of the trend in the average tar content of cigarettes and the percentage of females smoking (Scenario 9.1 and 9.2) indicate that this increase will continue until 2001, but by 2011 there is evidence of the beginning of a decrease in the age-standardised incidence rate (Figure 9.6). However, all other scenarios show a continuing increase in the incidence of female lung cancer. This is particularly marked for Scenarios 9.7 and 9.8 where the increase between 1991 and 2011 in age-standardised rates is projected to be of the order of 40%, and in the total number of cases is 62%, from 691 to 1120.



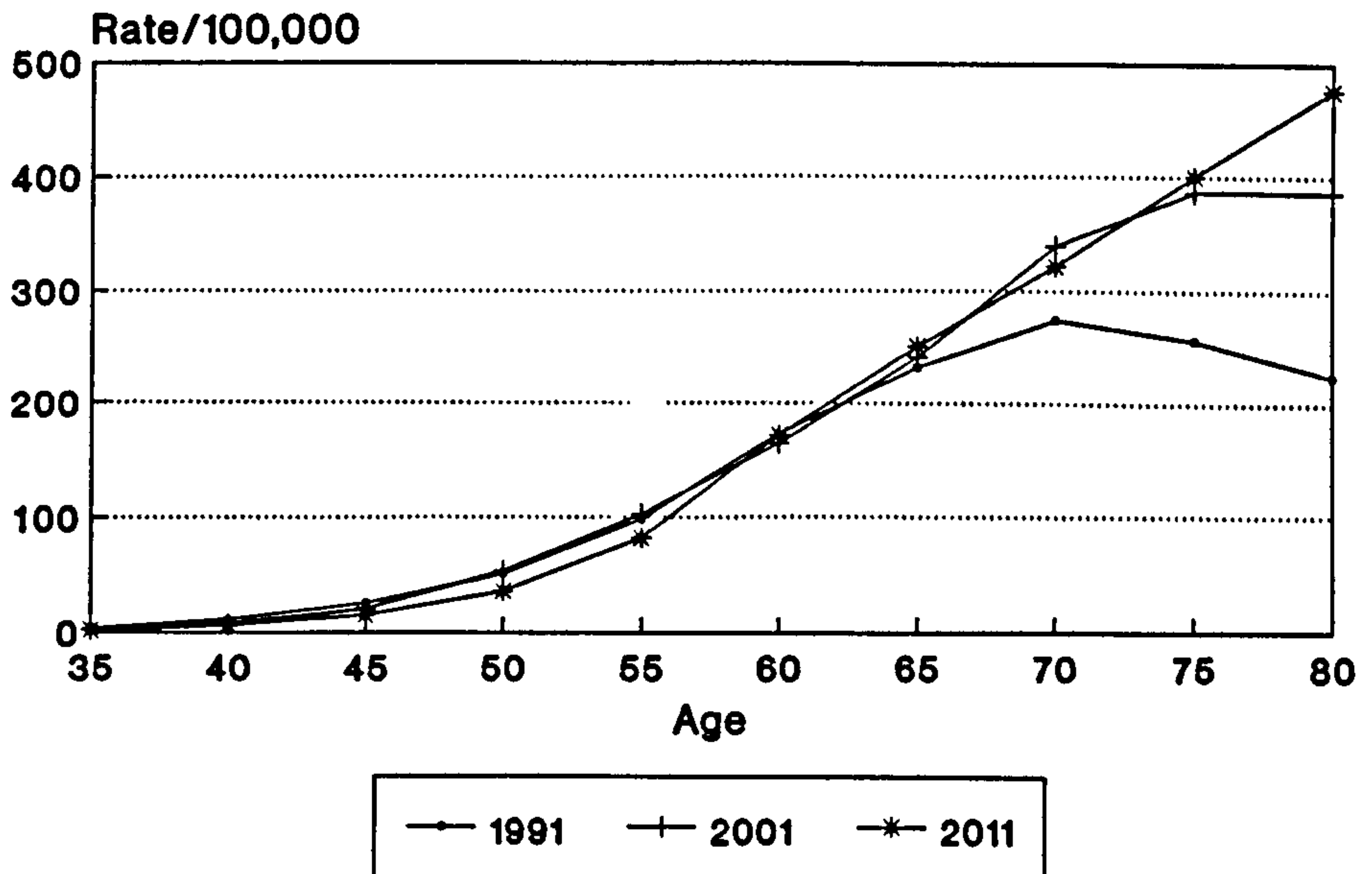
**Figure 9.7a Projected Lung Cancer Incidence Rates Scenario 9.1, Female**



**Figure 9.7b Projected Lung Cancer Incidence Rates Scenario 9.3, Female**

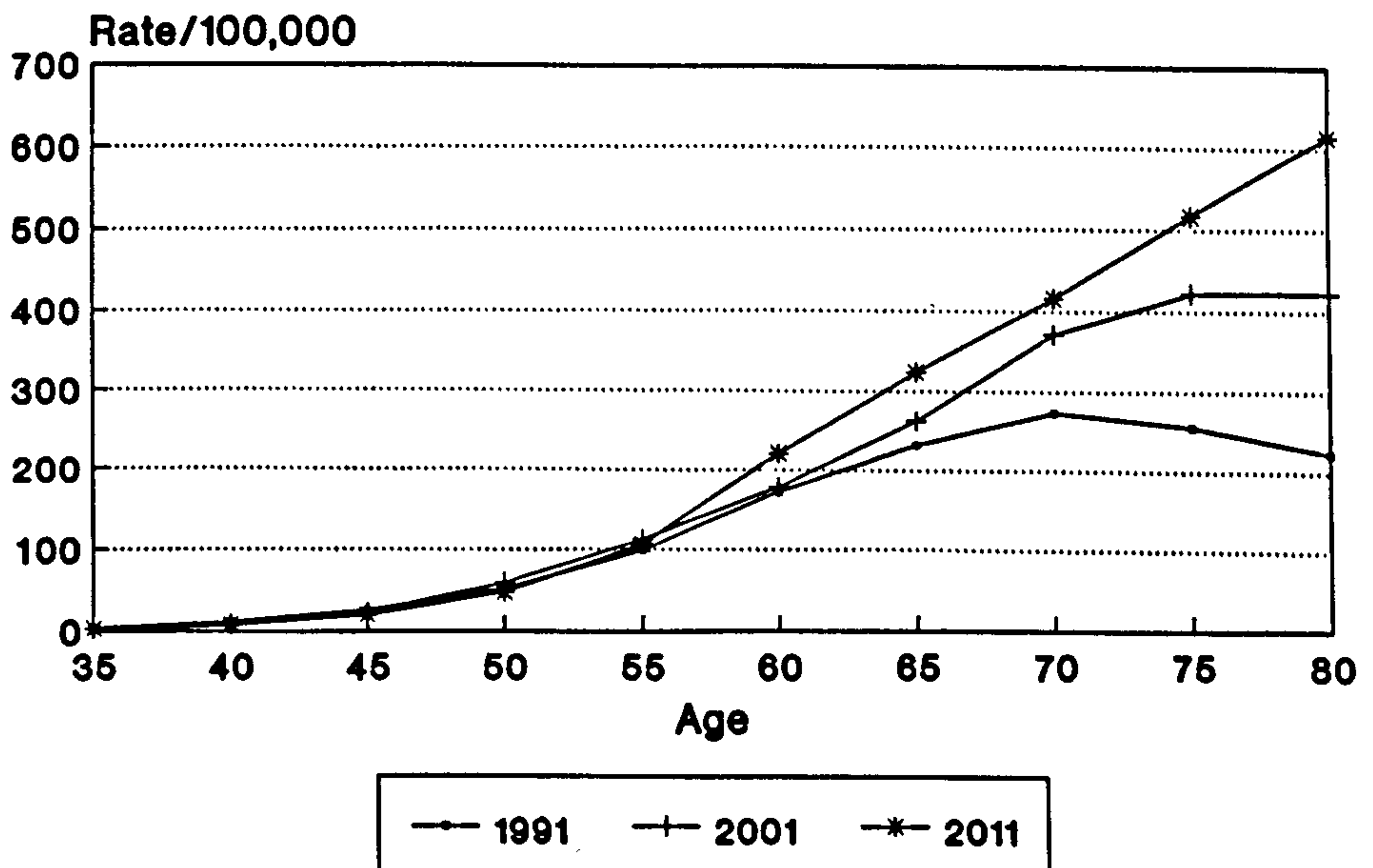


**Figure 9.7c Projected Lung Cancer Incidence Rates Scenario 9.5, Female**



Mersey Regional Cancer Registry

**Figure 9.7d Projected Lung Cancer Incidence Rates Scenario 9.7, Female**

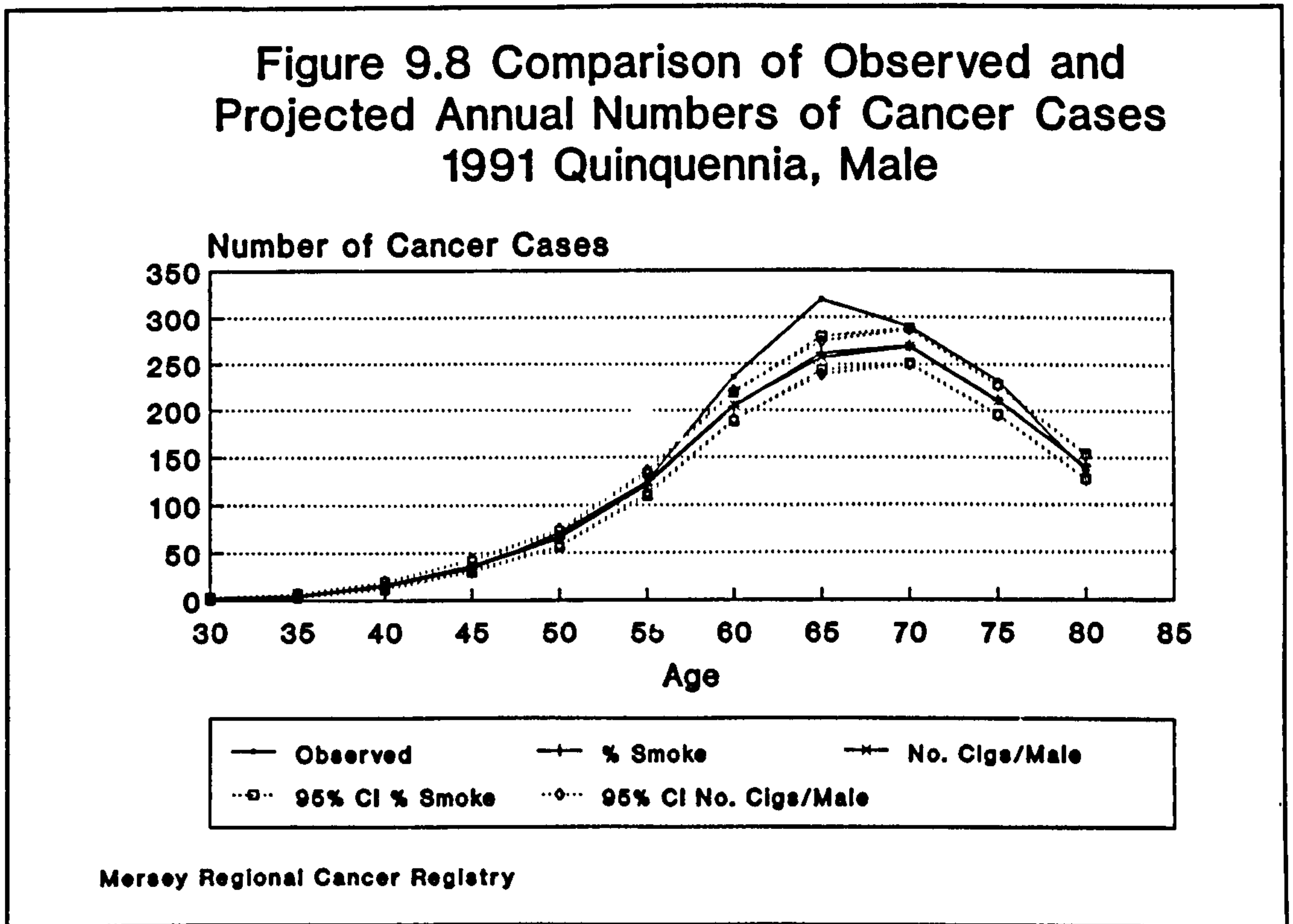


Mersey Regional Cancer Registry



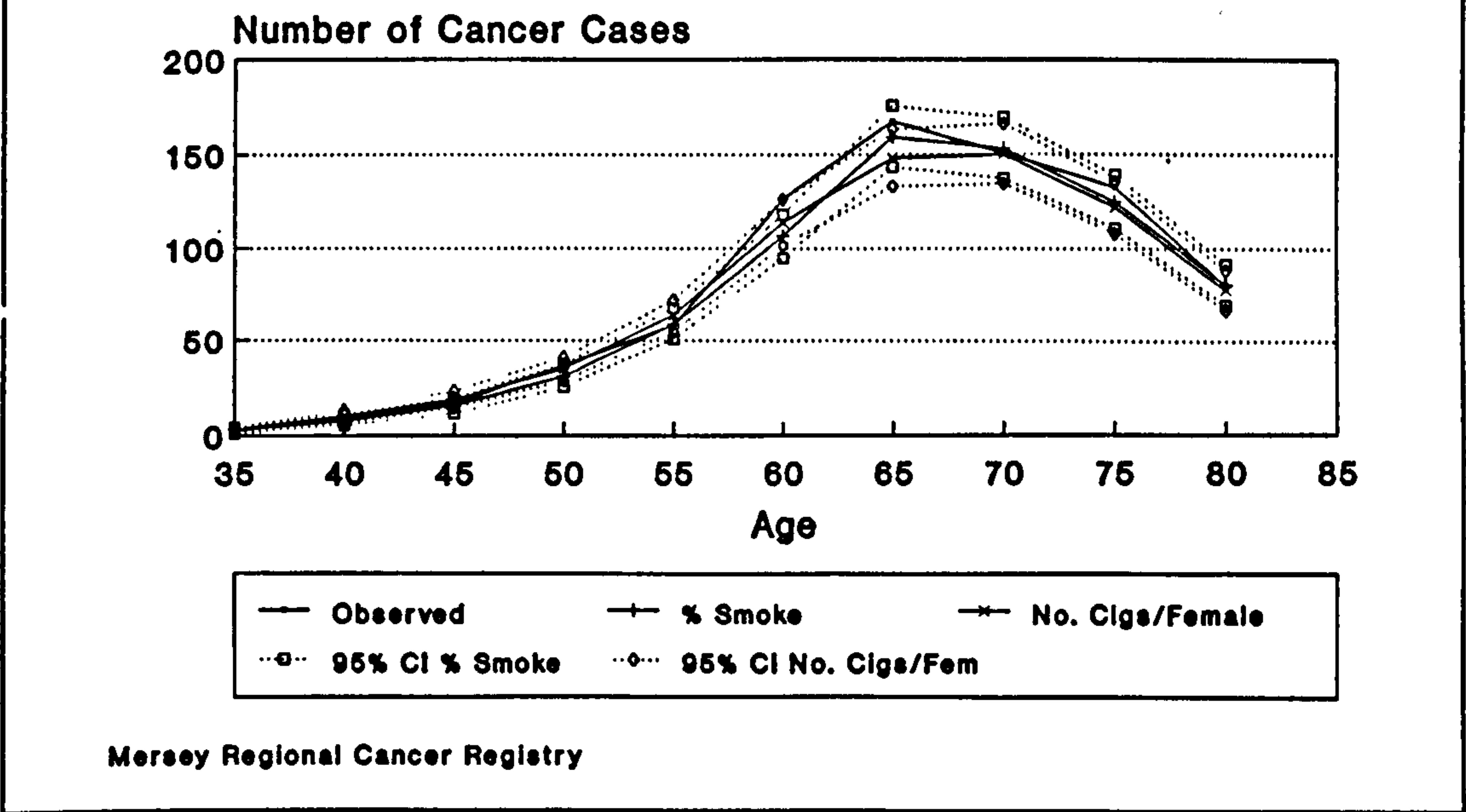
The age-specific curves for the years 1991, 2001 and 2011 show a pattern similar to that seen for the males for the quinquennia centred on 1971 through 1981 (Figures 9.7a, 9.7b, 9.7c, 9.7d and Appendix IV(2)). The age-specific curve for 1991 indicates a maximum age-specific rate for the age group 70-74 years, while for 2001 the age-specific incidence rate is maximum for the age group 75-80 years, and in 2001 the age specific rates continue to increase over the whole age range. Also of note is that for all scenarios the age-specific rates tend to decrease over time for the younger age groups. This decrease is apparent up to the age group 60-64 years for scenarios based on the percentage of females who smoke (9.1 to 9.4). For scenarios 9.5 to 9.8 the decrease in rates is only apparent up to the age group 55-64 years. This is similar to the pattern observed using age-cohort analysis (Figure 8.4). If females were to achieve the Health of the Nation targets for smoking, the age-specific incidence rates in the younger age groups are reduced. Since the number of cases occurring in these age groups is small, the overall effect on incidence is limited. Thus, only the results from the continuation of the present trends in smoking behaviour are presented. The effect of maintaining the 1986 level of average tar content of cigarettes is more marked for the scenarios including the average number of cigarettes smoked per female than for those including the percentage of females who smoke (Figure 9.7a, 9.7b, 9.7c and 9.7d). This

difference is considerably greater than that for males.



When the projected age-specific frequencies are compared to those actually observed in 1991, the overall pattern is similar (Figures 9.8 and 9.9). However both the models for males significantly underestimate the number of lung cancer cases occurring in the 60-64 years age group through the 75-79 year age group. For females the observed number of lung cancer cases fall within the 95 % confidence limits for the model based on the percentage of females who smoke. For the model based on the number of cigarettes smoked per female the observed number of lung cancer cases occurring in the 60-64 year and the 65-69 year age group exceeds the upper 95% confidence interval.

**Figure 9.9 Comparison of Observed and Projected Annual Numbers of Cancer Case 1991 Quinquennia, Female**



### 9.6 Health of the Nation Targets

Reduction of lung cancer incidence in males (30%) will achieve the Health of the Nation target (30%) if the average tar content of cigarettes continues its present rate of decline (Table 9.15). The effect of the average tar content maintaining its 1986 level is to decrease the reduction in lung cancer incidence to 17% for the Scenarios 9.1 to 9.4 and to 11% for the Scenarios 9.5 to 9.8. The attainment of the Health of the Nation targets for smoking does decrease the projected incidence rates in the younger age groups. However, since the incidence rates in these age groups are low, achievement of the targets for smoking does not affect the achievement of



**Table 9.15**      **Projected Percentage Reduction in Male Lung Cancer 1991 to 2011 for Males aged less than 75 years**

<b>Scenario</b>	<b>Age Standardised Rate<sup>1</sup> 1991<sup>2</sup></b>	<b>2011</b>	<b>% Reduction</b>
9.1	161.3	101.8	37%
9.2	161.3	100.8	38%
9.3	161.3	133.3	17%
9.4	161.3	134.6	17%
9.5	161.3	106.5	34%
9.6	161.3	106.1	34%
9.7	161.3	142.9	11%
9.8	161.3	142.3	12%

1. Standardised to the 1991 Mersey Region Population

2. Observed Rate

the target for lung cancer incidence.

The Health of the Nation Target for the reduction in lung cancer incidence in females under the age of 75 years (15%) will be met for Scenarios 9.1 and 9.2 (Table 9.16). A decrease in lung cancer incidence is also projected for Scenarios 9.3, 9.4, 9.5 and 9.6. In each of these scenarios the decrease is small. An increase in lung cancer incidence of 27% is projected for Scenarios 9.7 and 9.8; those incorporating no decline in the average tar content of cigarettes after 1986 and the number of cigarettes smoked per female. The apparent discrepancy



between these results and the overall trends reported above is due to the restricted age group involved in these comparisons, 35-74 years. For those females aged less than 60 years the lung cancer incidence is projected to decrease for all scenarios. The marked increases in lung cancer incidence occurs only in those aged over 70 years..

**Table 9.16**

**Projected Percentage Reduction in Female Lung Cancer 1991 to 2011 for Females aged less than 75 years**

<b>Scenario</b>	<b>Age Standardised Rate<sup>1</sup> 1991<sup>2</sup></b>	<b>2011</b>	<b>% Reduction</b>
9.1	99.3	83.0	16%
9.2	99.3	81.8	18%
9.3	99.3	95.7	4%
9.4	99.3	94.4	4%
9.5	99.3	98.3	1%
9.6	99.3	97.5	1%
9.7	99.3	126.8	-28%
9.8	99.3	125.8	-27%

1. Standardised to the 1991 Mersey Region Population
2. Observed Rates

### **9.7 Service Needs**

Between 1983 and 1987, an average of 150 patients received surgery, 135 received chemotherapy and 272 received radiotherapy each year for the treatment of lung cancer. If the trend in the average tar content continues

then service needs for male lung cancer will decrease by about 25% (Table 9.17). However, if the average tar content remains at the 1986 level the services needed will decrease by approximately 15%. The effects of attaining the Health of the Nation targets for the percentage of males smoking reduces the estimated service needs by only one or two cases, and the results are not presented here. The service needs for the scenario using the average number of cigarettes smoked per male are very similar to those for the percentage of males smoking. For the models based on the percentage of females who smoke (9.1 through 9.4) the annual services needs for the treatment of female lung cancer patients increases marginally from the quinquennia centred on 1985 to that centred on 2001 (Table 9.17). The greatest increase for these scenarios was for radiotherapy. For the models based on the number of cigarettes smoked per female the projected increase in service needs ranged from 16% for surgery to 28% for radiotherapy. These increases are not as marked as the overall increases in the age-standardised rates, but reflect the changes in the patterns of age-specific rates and the age-structure of the population, over this period.

Table 9.17

## Projected Service Needs for Incident Male Lung Cancer in Mersey Region

Quin- quennia	Scenario	Age	Surgery	Chemo- therapy	Radio- therapy
1985	Observed	30-54	23	28	34
		55-64	65	48	96
		65-74	56	44	110
		75-84	6	15	32
		Total	150	135	272
2001	9.1	30-54	21	26	31
		55-64	43	32	64
		65-74	43	33	88
		75-84	5	13	28
		Total	112	104	211
2001	9.3	30-54	23	28	34
		55-64	47	35	70
		65-74	47	36	97
		75-84	5	14	30
		Total	122	113	231
2001	9.5	30-54	22	27	33
		55-64	45	33	66
		65-74	43	34	90
		75-84	5	13	27
		Total	115	107	216
2001	9.7	30-54	25	30	36
		55-64	49	37	73
		65-74	48	37	99
		75-84	6	14	30
		Total	128	118	238

Table 9.18

## Projected Service Needs for Incident Female Lung Cancer in Mersey Region

Quin- quennia	Scenario	Age	Surgery	Chemo- therapy	Radio- therapy
1983-87	Observed	35-54	13	14	16
		55-64	26	30	45
		65-74	20	22	46
		75-84	1	6	10
		<b>Total</b>	<b>60</b>	<b>72</b>	<b>117</b>
2001	9.1	35-54	13	12	16
		55-64	20	23	35
		65-74	26	29	61
		75-84	3	11	21
		<b>Total</b>	<b>62</b>	<b>76</b>	<b>133</b>
2001	9.3	35-54	13	14	16
		55-64	21	24	37
		65-74	27	30	63
		75-84	3	12	22
		<b>Total</b>	<b>64</b>	<b>80</b>	<b>138</b>
2001	9.5	35-54	15	16	19
		55-64	24	28	41
		65-74	28	30	64
		75-84	3	10	20
		<b>Total</b>	<b>70</b>	<b>84</b>	<b>144</b>
2001	9.7	35-54	17	18	20
		55-64	26	30	45
		65-74	30	33	70
		75-84	3	11	21
		<b>Total</b>	<b>76</b>	<b>92</b>	<b>156</b>



## 9.8 Summary of Projections

This chapter has investigated projections based on the adaptations of the age-period-cohort model. In the previous chapter, the projections were based on linear extrapolations of the cohort and period parameters. In this chapter smoking variables have been included to assist in the extrapolation.

The smoking variables that are investigated are:

- the average tar content of cigarettes
- the percentage of the population who smoke, for each sex separately
- the number of cigarettes smoked per person, for each sex separately

The average tar content of cigarettes is considered to be a period effect, while the percentage of the population who smoke and the average number of cigarettes smoked per person are considered to be cohort effects.

The average tar content of cigarettes contributed significantly to the models for both males and females. A lag period of 10 years between trends in the average tar content and trends in lung cancer incidence gave the best fit for both males and females. Two set of models were investigated; one including average tar content and the percentage of the population who smoke, and the other

including the average tar content and the average number of cigarettes smoked per person. If both the cohort smoking variables were included in the models concurrently, the coefficient for the number of cigarettes smoked per male is negative and thus non-sensical, and for females the number of cigarettes smoked per female did not contribute significantly to that including the percentage of females smoking. The final squared deviance for the models for males were larger than that obtained for the age-period-cohort analysis. For the females the final scaled deviances were approximately equal. However, the models including information on smoking behaviour allow investigation of the effect of changes in smoking behaviour.

Two assumptions for the future level of the average tar content of cigarettes are investigated: viz,

- t1> estimates based on extrapolation of the log linear trend in tar content to 1991 and 2001,
- t2> estimates based on assuming that the tar level remains constant after 1987

Two assumptions for the future percentage of the population who smoke and the number of cigarettes smoked per person are also investigated: viz,

- c1> to estimate the values of the smoking variables by linear extrapolation of the log of the smoking variables
- c2> to assume that the more recent birth cohorts will achieve the stated goal for the Health of the Nation; viz 20% smoking and a reduction of 40% in the number of cigarettes smoked.

These two sets of assumptions give rise to 8 different projection scenarios (Table 9.11).

The projections based on these 8 scenarios indicate that male lung cancer incidence will continue to decline for all age groups and for all scenarios. The age-specific incidence for female lung cancer will decline for the younger females (aged less than 70 years for scenarios 9.1 to 9.4, and those aged less than 60 years for scenarios 9.5 to 9.8). For the older females the age-specific incidence rates will increase dramatically.

The projections for 1991 do not agree as closely with the rates actually observed as those projections based on the age-period-cohort analysis. However, the 95% confidence intervals for these models are much narrower than those based on the age-period-cohort analysis. The difference in rates due to attaining the Health of the Nation



targets for smoking is small. This is in part due to the form of the model that does not allow variation in the smoking variables for the cohorts prior to 1954-63. The effect of the different assumption regarding the average tar content of cigarettes is more marked with the decrease in age-specific rates being much less for the scenarios where the average tar content does not decrease.

Service needs are projected to decrease for males, from the quinquennia centred on 1985 to that centred on 2001. For the females there is a projected increase in service needs for females. For the worst case scenario there will be no decrease in service needs. The worse case scenario are those based on constant average tar content of cigarettes and the number of cigarettes smoked per person (Scenarios 9.7 and 9.8). For the best case scenario, the reduction in service needs will be greater than 10% for all types of treatment. The best case scenarios are those based on a decline in the average tar content of cigarettes and the percentage of the population who smoke (Scenarios 9.1 and 9.2). The models based on the percentage of the population who smoke fit the data better, and therefore it is probable that there will be a decrease in service needs.



Chapter 10

**DISCUSSION**

The Health of the Nation strategy (Department of Health 1991) has set targets for the reduction of lung cancer mortality. Targets for the reduction in smoking have also been set to assist in the attainments of the targets for the reduction of lung cancer mortality. The Chief Medical Officer's Expert Advisory Group on Cancer (Department of Health 1994) gives guidelines for the organisation of cancer services. The implementation and evaluation of both would be assisted by accurate projections of future cancer incidence. Models to describe the trends in lung cancer incidence have been developed. Projected lung cancer incidence rates for the next 20 years have been calculated, based on the models. The projections have been used to determine whether the Health of the Nation targets will be met and to quantify the service needs for lung cancer patients by the year 2001. The results of this research are discussed in this chapter. The discussion will cover:

- 
- comparison of the models that are developed with particular reference to
    - extrapolation over time compared to inclusion of information on smoking behaviour
    - trends over periods compared to trends over cohorts
    - quality and availability of data on smoking behaviour
  - comparison of projected lung cancer incidence
    - does the inclusion of information on smoking behaviour improve the projections?
    - do the projections depend on the model used?
    - which is the preferred model for projections?
  - implications for the Health of the Nation strategy
    - will the targets for the Health of the Nation be met?
    - does attainment of the Health of the Nation targets for the reduction of smoking imply that the target for the reduction in lung cancer will be attained?
    - are the correct smoking indicators being monitored?
  - implications for Service Needs
    - do the projections of service needs depend on the model used?
    - what are the implications of the Expert Advisory Group on Cancer's guidelines?
-

## 10.1 Methods of Projection

In this thesis the trends in lung cancer incidence in Mersey Region are investigated. Models are developed to describe these trends and to project future lung cancer incidence. The projections are used to evaluate the possibility of attaining the Health of the Nation targets, and to determine future service needs for lung cancer patients in Mersey Region. The projections will also be useful in answering other questions relating to future lung cancer incidence.

The accuracy of the projections of the number of cases of cancer for a given year depends, not only on the accuracy of the projected rate, but also on the accuracy of the projected population figures. The population projections that are used in this thesis are those provided by OPCS (1995). These are based on the assumptions that local trends in births, deaths and net migration will continue for the period of projection. If there is a change in the economic situation in Mersey Region there may also be a change in the trends in births and net migration. However, it is beyond the scope of this thesis to investigate such changes.

The simplest type of model describes the trends in the age-standardised rates over calendar period, for males and females separately. Since 1974 the age-standardised



lung cancer incidence rates for males have decreased by about 1.3% per year, while in females they have increased by about 3.5% per year. The observed increase in females, obviously leads to concern. In fact, in the Liverpool District Health Authority, the age-standardised incidence rate for lung cancer in females has exceeded that for breast cancer (Williams et al 1993). While the age-standardised incidence rates give an overall picture of the trends in lung cancer incidence, they do not allow the investigation of trends among the age-specific rates. The broad-band age-specific incidence rates also decrease for the males and increase for the females. The decrease in the males aged 45-64 years (2.5% per year) is larger than that for the males age 65 years and over (0.4% per year). Younger females also fare better than older females, even though the trend for both is for an increase. The increase in lung cancer incidence in females age 45-64 years (2.3% per year) is only half that for those age 65 years and older (4.8% per year). This pattern may indicate that the younger population has been exposed to less cigarette smoke than the older population. As the low risk younger males get older, the rate of decrease in lung cancer incidence among males aged 65 years and older may accelerate. Similarly the rate of increase in females aged 65 years and older may decelerate. These changes can best be described as changes over birth cohorts. The models developed to

describe the age-standardised and the broad-band age-specific lung cancer incidence rates do not allow investigation of these types of changes, because the trends in lung cancer incidence are assumed to be dependant on calendar period alone. The advantages of extrapolation of trends over calendar period for either age-standardised or age-specific lung cancer incidence rates are that it is relatively simple to do, and is conceptually easy to understand.

The age-period-cohort analyses indicate that trends in lung cancer incidence in males are dependant on both calendar period effects and birth cohort effects, while the trends in females are dependant on birth cohort effects but not period effects. The decrease in males cannot be quantified for period and cohort effects independently because of the problem of non-identifiability of the parameter estimates. The model for females is based only on age and cohort. The result is somewhat surprising in that there is a 5% decrease in lung cancer incidence every 5 years for the 5 most recent birth cohort included in the analysis. The overall increase in female lung cancer incidence up to the present is due to increases in incidence in those cohort born between 1874-83 and 1928-37. The slowing down in the rate of increase in the female age-standardised incidence rates in recent years is due, in part, to the decreases in incidence in the more recent cohorts (Section 4.11.2).



The effect is not very marked because the females in these birth cohorts have not yet reached an age where lung cancer incidence is high. If this decrease in lung cancer incidence over cohorts were to continue, an overall decrease in lung cancer incidence among females would be expected in the future.

These models can be used for projections, on the assumption that the relationship between incidence and time can be extrapolated. For the age-standardised and the broad-band age-specific rates the log of the rates is linearly related to calendar period. Thus future incidence rates can be estimated through linear extrapolation of past trends. In order to calculate projected incidence rates from the age-period-cohort models, future period and cohort parameters are estimated through linear extrapolation of the more recent period and cohort parameter curves. This assumes that for the more recent periods and cohort, the relationships with time are essentially linear. However, if period and/or cohort effects are included in the model it is because these effects are essentially non linear. If the relationships were linear, the effects would be modelled using the drift parameter (Section 8.1). Over the whole period the cohort effects for lung cancer in MRHA were non-linear, but essentially linear for the more recent cohorts. A similar pattern appears for the period effects in the model for males.

To overcome these problems and to improve the accuracy of the projections, information on risk behaviour is included in the models. The main risk factor for lung cancer is cigarette smoking, accounting for over 30% of lung cancer incidence. Occupational exposure to various carcinogens also contribute to lung cancer incidence. Since, the effects of occupational exposure are confounded with the smoking behaviour, and account for only 10% to 20% of the lung cancer burden, this thesis considers only the effects of smoking behaviour on lung cancer incidence.

Table 10.1		Percentage of the Population who Smoke: A comparison of the MRHA and GB	
Year	Sex	MRHA	GB
1990	Male	30%	31%
1992	Male	29%	29%
1990	Female	31%	29%
1992	Female	28%	28%

GHS 1994 Table 10.1

The data on smoking behaviour that are included in the model should be population based and relevant to the population for which the projections are to be calculated. There is little information on cigarette smoking for Mersey Region. Although smoking is more common in the North West than in Great Britain as a whole, the trend in smoking behaviour mirrors the



national trend (Wald and Nicolaides-Bouman 1991, Table 10.2). The percentages of males and females who smoke in MRHA for 1990 and 1992 are comparable with the corresponding figures for Great Britain (General Household Survey 1994). These are the only historic smoking data available for MRHA apart from some ad-hoc surveys. Thus national figures have been used to investigate the trends in lung cancer incidence in Mersey Region. The inclusion of these smoking data improves the accuracy of the projections and the ability to investigate changes in trends of smoking behaviour on local lung cancer incidence. The methods described in this thesis could be used by other registries without the necessity of having local smoking information. However, it would be necessary to determine that the local smoking trends were not radically different from the national average.

In England the TAC has conducted annual surveys on smoking behaviour in the United Kingdom since 1948, which they published. Since 1988 the data from these surveys have not been made available (Wald and Nicolaides-Bouman 1991). The General Household Survey (GHS) has also published data on smoking behaviour since 1972 (OPCS 1975-1990). The two sources of data are not completely equivalent because the definition of a smoker and the questions asked relating to quantity smoked are different. The TAC data on quantity smoked are adjusted

to take into account the known sales to the public, to overcome the potential under-reporting of quantity smoked (Wald and Nicolaides-Bouman 1991). Data from the GHS are not similarly adjusted. Thus the data from the GHS would indicate a lower percentage of the population smoking and fewer cigarettes smoked than the TAC data. The data that are used in this thesis are all derived from TAC surveys so that consistent definitions are used. If more recent information on smoking behaviour are to be included in the models, there are potential problems with the discontinuity of the smoking data.

On an individual basis the risk of lung cancer is dependant on the quantity smoked, the duration of smoking, the age at starting to smoke, the composition of the cigarettes smoked and if quit the time since quitting. At a population level, smoking variables used to explain trends in lung cancer incidence cannot be deemed to be measuring a causative relationship. They are more indicators of trend. Data on the quantity smoked, the percentage of the population smoking and the composition of cigarettes are available and have been included in the analyses. It is believed that smoking habits are acquired early in life (Hammond 1966) and vary with birth cohort (Doll and Hill 1964). Thus it can be assumed that quantity smoked and the percentage of the population who smoke would be determined by birth cohort and can thus be considered to be a cohort effect. Smoking

behaviour usually begins in the late teens or early 20s, and stabilises by the mid 20s. The data for the age group 25-34 years are used as the measure of each cohorts exposure to cigarette smoking. The use of this age group as an indicator for cohort exposure to cigarette smoking does not allow for changes in smoking habits within cohorts.

Effective anti-smoking campaigns may lead to cohorts changing their smoking habits differentially. Age-specific data on the percentage of the population who had quit smoking and the average time since they had quit, would assist in the evaluation of the effects of quitting smoking on lung cancer incidence. Adequate published data are not available. Thus it is not possible to investigate the effects of Health education messages to quit smoking on population incidence rates.

The average tar content is included as a period effect because changes in the composition of cigarettes are thought to affect all smokers at the same time. The average tar content of cigarettes is calculated using sales figures for the individual brands. If brand allegiance is determined near the beginning of the smoking habit, then changes in tar content of cigarettes smoked may not be the same for all age groups. The available data does not allow this problem to be addressed.



The smoking data that are used in the development of the models are published data, with the exception of some of the age-specific data for quantity smoked and the percentage of the population smoking. In these cases age-groups had to be combined because of the changing age groups for the published data (Table 3.5). Other models that have been developed to incorporate smoking information have estimated historic data from cross sectional surveys (Brown and Kessler 1988), or back generated the data assuming no changes in smoking patterns (Stevens and Moolgavkar 1979, 1984). Models developed on risk factor data that have been extrapolated or interpolated from published data are dependant on the assumptions used in the estimation of the risk factor data. If these assumptions are incorrect, the models may also be incorrect.

In the trend analysis of both the age-standardised rates and the broad-band age specific rates, the quantity smoked and the percentage of the population who smoke have been included as period effect. The lag periods between changes in smoking behaviour and changes in lung cancer incidence have to be estimated. In other analysis the lag periods have been determined through correlation analysis for each variable separately (Hakama and Pukkala 1984). Cigarette smoke is thought to be both an early stage and late stage carcinogen (Hayes and Vineis 1989). Two distinct lag periods would then be appropriate



(Hakama and Pukkala 1984). For simplicity it is assumed in this thesis that the lag period for any given model would be the same for all smoking variables. The lag periods are determined concurrently with the determination of the best predictive smoking variables for lung cancer incidence. Lag times of 10 years, 15 years, 20 years and 25 years are investigated. The variation of both the lung cancer incidence data and the smoking data about the general trends would make it difficult to discriminate among individual years of lag periods. The assumptions that all lag period are the same for all smoking variable may explain why it is difficult to determine the lag times between smoking behaviour and lung cancer incidence for these models (Sections 6.2 and 7.2). If the information on quantity smoked and the percentage of the population who smoke are included as cohort effects the problem of lag time estimation is overcome. The smoking behaviour is assigned to the relevant birth cohort.

In the analyses relating age-standardised and broad-band age-specific rates to smoking variables, the coefficients for the average tar content of cigarettes and/or the percentage of smokers who smoke plain cigarettes were negative. These models were discounted as being implausible. For the females the trends in the lung cancer incidence rates are essentially linear and increasing over the period of consideration. The trends

in the average tar content of cigarettes and the percentage of smokers who smoke plain cigarettes are also approximately linear, but decreasing. Thus there would be a strong negative correlation between lung cancer incidence and these two measures of the composition of cigarettes. The relationship is discounted as implausible because the literature has shown that lower levels of tar in cigarettes are associated with lower risks of lung cancer (Lubin 1984a, 1984b). Other complications may have produced an implausible inverse relationship. The inverse relationship may also be due to the inclusion of the variables relating to quantity smoked as period effects. The negative coefficients for the composition of cigarettes correcting for the overestimation of the effects of reducing quantity smoked.

The average tar content of cigarettes contributes significantly to most of the models that fit male lung cancer incidence. Apart from the models for the broad-band age-specific rates, the coefficient of the log of the average tar content is approximately equal to 0.7, varying between 0.61 for the 15 year lag for the age-standardised rates and 0.97 for the 20 year lag for the age-standardised rates (Table 6.3). Thus for each percentage decrease in the average tar content there is approximately a 0.7% decrease in lung cancer incidence. This consistency in the parameter estimates for the average tar content of cigarettes increases our

confidence in the models.

The age-period-cohort model would indicate that trends in female lung cancer incidence are dependant only on age and cohort effects. When information on smoking behaviour is included in the model the percentage of females who smoke or the number of cigarettes smoked per female contributes significantly as a cohort effect and the average tar content of cigarettes contributes significantly as a period effect. The average tar content of cigarettes may be reflecting changes in total exposure to cigarette smoke not explained by the cohort variable. For the model that includes the percentage of females who smoke the coefficient for the average tar content of cigarettes is approximately half that for the males. Females have smoked lower tar cigarettes than males and thus changes in the sales adjusted average tar content of cigarettes may reflect changes in male smoking habits more than those of females. Thus the effect of changes in the average tar content would be smaller in females than in males. In the model including the number of cigarettes smoked per female the coefficient for the average tar content of cigarettes is similar to that for males. Brown and Kessler (1984) in the United States found the coefficient for the total tar consumption for females was larger than than for males. However their tar measure also included an adjustment for the quantity smoked, and this may explain the differences in the results.



The coefficients for the number of cigarettes smoked per person are also broadly consistent between models for each sex. The male coefficient is approximately half that of the female coefficient. Similarly the coefficient for the percentage of males who smoke is approximately half that of the coefficient for the percentage of females who smoke. This may also be due to the difference in smoking patterns of males and females, with particular reference to patterns of quitting and the distributions of the quantity smoked amongst the smokers. Since there is no adequate published data on quitting or the distribution of quantity smoked it is not possible in this thesis to investigate these differences.

## 10.2 Comparison of Projections

The projections based on the simplest models, linear extrapolation of the age-standardised and broad-band age-specific incidence rates over calendar period give higher estimates of lung cancer incidence than those based on similar models including information on smoking behaviour. This is not unexpected because both the average tar content of cigarettes and the quantity smoked began to decline rapidly in the early 1970s. These models have assumed that changes in smoking behaviour would affect all age groups in the same way. However, while average tar is considered to be a period effect, the quantity smoked and the percentage of the population who



smoke are considered to be cohort effects. Thus the decreases in quantity smoked and the percentage of the population who smoke are likely to affect the younger age groups more than the older ones. The patterns of smoking are more fixed in the older age groups. Therefore the projections based on models including quantity smoked or the percentage of the population who smoke are likely to be too low.

The projections based on linear extrapolation of the period and/or cohort parameter estimates from age-period-cohort analysis are smaller than those based on similar models incorporating information on smoking behaviour. The one exception to this is for Model 9.1 for females which includes linear extrapolation in the trends in the average tar content of cigarettes and the percentage of females who smoke. The differences between the models based on linear extrapolation of trends in the period and cohort parameter estimates and those including information on smoking behaviour may be due to problems with the assumptions relating to future smoking behaviour, or they may be due to the assumption that the smoking behaviour in the age group 25-34 years can be used as a measure for the smoking behaviour of the birth cohort. As with the models based on the age-standardised and age-specific rates, information on quitting smoking would help address these problems.

Changing the assumptions related to the future values for the percentage of the population who smoke or the number of cigarettes smoked per person has minimal effect on the projections (Section 9.5.4). The model assumes that changes in smoking behaviour will only affect cohorts born after 1959. For the period of projection employed here these cohorts would be less than 50 years old. In this age range the number of incident lung cancer cases per year is small relative to the total number. If it were possible to include information on quitting and/or starting smoking in the models it may be possible to determine the effects of changes in smoking behaviour for the whole population.

The assumptions relating to the sales adjusted average tar content of cigarettes have a more marked effect on the projections. This is because changes in the average tar content of cigarettes is assumed to affect all age groups in the same way. Unfortunately, this measure of the composition of cigarettes is no longer available because the tables on the market share of brands of manufactured cigarettes in the UK are no longer published. If the recent decline in the average tar content of cigarettes does not continue at least an extra 325 males and 120 females will develop lung cancer in 2011.

The confidence intervals for the projected number of

incident lung cancer cases for each 5 year age group based on age-period-cohort analysis include the observed number of cases for 1991. The confidence intervals for the projections based on models including information on smoking behaviour are much narrower, but for males do not include the observed numbers of cases for the age groups 60-64 years and 65-69 years. Whether this is due to a poor fit of the model, or is due to some change in registration practise at the registry is difficult to determine.

The projection for the age-specific rates for male lung cancer incidence show a steady decline for all age groups over the period of projection (Figure 9.5). The pattern for females is more interesting and mirrors that for males approximately 30 years earlier. In 1991 females in the 65-69 year age group had the highest age-specific lung cancer incidence rate (Figure 9.7). By the year 2011 the age-specific incidence rates are projected to increase with age across all the age groups. Also of note is the decrease over calendar period in the age groups up to 65-69 years between 1991 and 2011. The incidence rates for the age groups 70-74 years and 75-79 years in 2011 are projected to be less than those for the same age groups in 2001.

These patterns are reflecting the pattern of exposure to cigarettes among cohorts. Smoking among females did not



start to become popular until World War II although some females had started to smoke in the 1920s. Those females who were over 70 in 1991 would have been born before 1921 and be over the age of 25 years by the end of the war. They would not have had the same exposure to cigarette smoke in their younger years as those females who were born after 1921. As the cohort of females who have been more exposed to cigarette smoke (born after 1921), becomes older the age-specific lung cancer incidence rates in those aged 70 years and above can be expected to increase. With the decrease in smoking in more recent years the age-specific lung cancer incidence rates will start to decline in the younger age groups. In the projections up to 2011 the pattern in the older age groups reflects the start of the smoking epidemic. Hopefully the pattern in the younger age groups reflects the beginning of the end of the smoking epidemic.

A similar pattern was seen among the males starting in 1971. Males started smoking manufactured cigarettes in 1895, 25 years before females. The habit had become popular among males by World War I in 1914. Therefore, given the decreases in lung cancer incidence seen in males, and the similarities in the decrease in smoking patterns in both males and females, it would be expected that lung cancer incidence in females will begin to decline across all age groups.



Given the variations in the projections based on the different models considered in this thesis, is it possible to determine which is the preferred model? Trends in lung cancer incidence appear to depend on variation over cohort. Therefore the model should include cohort effects. The models based on age-standardised and broad-band age-specific incidence rate that were considered in this thesis will not explain the trends as accurately as the models based on age-period-cohort analysis. Among the age-period-cohort models those including information on smoking behaviour give estimates of the age-specific rates that have considerably smaller 95% confidence intervals. However, when the projected age-specific incidence rates for 1991 are compared with those actually observed, the rates for the age groups 60-64 years and 65-69 years, fall outside the 95% confidence intervals. Inclusion of the percentage of the population smoking gives a better fit than inclusion of the number of cigarettes smoked per person, and thus the projections based on the percentage of the population who smoke should be more accurate. The average tar content of cigarettes is included in both age-period-cohort models. Trends in this variable have a marked effect on the projections. Thus, if trends in smoking habits are to be investigated, the preferred model would be one based on age-period-cohort models with the inclusion of the average tar content of cigarettes as a period effect and the percentage of the population who smoke as a cohort

effect. If data on these smoking variables are not available, then projections based on the age-period-cohort model would be preferred over those based on the age-standardised rates or broad-band age-specific rates for lung cancer incidence.

### 10.3 Implications for Health of the Nation

In the 1992 strategy for the Health of the Nation lung cancer was one of the targeted diseases (Department of Health 1991). The target for the reduction of lung cancer is:

To reduce the death rate for lung cancer under the age of 75 by at least 30% in men and by at least 15% in women by 2010 (from 60 per 100,000 for men and 24.1 per 100,000 for women in 1990 to no more than 42 and 20.5 respectively)

This target is expressed as a reduction in mortality. Between 1986 and 1990 in Mersey Region the mortality for lung cancer was 75.4 per 100,000 for males and 32.4 per 100,000 for females. Thus to achieve the targets set in the Health of the Nation, there would need to be a 44% reduction in mortality in males and a 37% reduction for females in Mersey Region. The main thrust of the Health of the Nation is to prevent lung cancer occurring and thus reduce incidence; not to improve survival. The median survival time for lung cancer in Mersey Region is 4 months, with about 10% of cases surviving 2 years. The

incidence-mortality ratio for lung cancer in Mersey Region in 1989 was 1.12 in males and 1.11 in females (OPCS 1994). Therefore, the trends in incidence can be expected to be similar to those for mortality. If the targets of the strategy are to be met a decline in incidence would be required similar to that for mortality. In this thesis, there has been no adjustment in the targets set by the Health of the Nation strategy for MRHA.

At least 80% of lung cancer is attributable to smoking. Therefore one mechanism to reduce the burden of lung cancer is to reduce smoking among the population. In the Health of the Nation strategy, targets have been set for the reduction of smoking. These are:

1. To reduce the prevalence of cigarette smoking to no more than 20% by the year 2000 in both men and women (a reduction of at least 35% in man and 29% in women, from a prevalence in 1990 of 31% and 28% respectively)
2. To reduce consumption of cigarettes by at least 40% by the year 2000 (from 98 billion manufactured cigarettes per year in 1990 to 59 bn)
3. In addition to the overall reduction in prevalence, at least 33% of women smokers to stop smoking at the start of their pregnancy by the year 2000
4. To reduce smoking prevalence of 11-15 year olds by at least 33% by 1994 (from about 8% in 1988 to less than 6%)



This thesis investigates the relationship between lung cancer incidence, and the percentage of the population who smoke and the quantity smoked. The relationship between the trends in lung cancer and the indicators relating to smoking in pregnancy and smoking in childhood have not been addressed.

Quitting smoking for the duration of pregnancy is beneficial for the foetus, but if the mother then returns to smoking after the pregnancy, there will be little effect on her risk of lung cancer. Data on smoking in pregnancy is available from several studies and would indicate that less than one quarter of female smokers stop smoking during pregnancy. These studies were carried out between 1958 and 1984 and cover various regions in Great Britain (Wald and Nicolaidis-Bouman 1991). There appears to be no evidence for a trend over time in the percentage of females who quit smoking for the duration of their pregnancies, assuming there is no geographic variation. Therefore inclusion of information of the quitting rates during pregnancy is unlikely to contribute to the models for the trends in lung cancer incidence.

Those people aged 11-15 years at the beginning of the 21st century, will not have reached the age when lung cancer is likely to be a threat by 2011. Data on the smoking habits for children are also available from ad-

hoc studies from 1966 to 1988. As with the data on pregnant women, the studies cover different areas of Great Britain, and no obvious trend over time is evident. Therefore inclusion of this variable is unlikely to contribute to the model. Thus the achievement of targets 3 and 4 are unlikely to affect the projected cancer incidence in 2011.

The two other targets relate to the percentage of the population smoking and to the quantity of cigarettes smoked. Age-specific data on these two indicators are available on an annual basis from 1948 onwards. On an individual basis the risk of lung cancer is related to the amount smoked, usually measured as number of cigarettes smoked per unit time. With population based data, it is possible to include this information as either cigarettes smoked per person, or cigarettes smoked per smoker. The number of cigarettes smoked per smoker measures the quantity smoked by smokers, but does not include any information on the percentage of the population smoking, or give any information of the distribution of quantity smoked among smokers. Heavy smokers are at a higher risk of lung cancer than light smokers. Therefore if smokers are smoking more cigarettes per unit time then the incidence could be expected to rise, if the percent of the population smoking stayed the same. However, if the percent of the population smoking increases over time and the number of

cigarettes smoked per smoker remained the same the incidence would be also be expected to rise. The number of cigarettes smoked per person includes information on the percent of the population smoking and the number of cigarettes smoked per smoker. Trends in this variable could therefore be expected to reflect more closely trends in lung cancer incidence. However, this variable would not differentiate between a small percentage of the population smoking a large amount and a large percent of the population smoking a small amount. Whether this has an effect on the population risk needs to be explored.

The target for the reduction in quantity smoked is given as the total consumption of cigarettes for the whole population. The required reduction is 40%. In translating this into targets for the reduction of cigarettes smoked per person, the current levels for both males and females were reduced by 40%. It may be more effective for the reduction in lung cancer incidence if these targets had been apportioned in some other fashion. The Health of the Nation has no targets for the composition of cigarettes in relation to tar and other compounds. The percent of cigarettes smoked that are plain is now close to zero for both males and females, and therefore cannot be reduced further.

For most models developed in this thesis the Health of the Nation targets for male lung cancer will be met, or



nearly met, if the present trends in lung cancer incidence continue. The only exception to this is for males aged 65 years and over. These projections assume that trends in smoking behaviour will continue for the period of projection. If the current trend in the reduction of the percentage of males who smoke and the number of cigarettes smoked per male continue the Health of the Nation targets for smoking behaviour will not be met by the year 2000 (Table 9.13). Thus the targets for lung cancer incidence in males will be met even if the targets for smoking behaviour were not met.

Extrapolation of the current trends for females over calendar period would indicate that the Health of the Nation targets will not be met. In fact lung cancer incidence is projected to at least double for the age-standardised rates and the age-specific rate for females aged 65 years and older. The trends in female lung cancer are dominated by cohort effects. If the changes in female age-specific lung cancer incidence are modelled as trends over birth cohorts rather than calendar periods, then the age-standardised rates for females age less than 75 years are projected to decrease by 10%.

If information on smoking behaviour is included as period effects in the models for age-standardised and broad-band age-specific rates, the reduction in lung cancer incidence is probably over estimated. Thus the

projections that the reductions in age-standardised rates for females will meet their targets is probably erroneous.

The projections based on age-period-cohort analysis that also include information on smoking behaviour are most affected by the assumptions relating to the average tar content of cigarettes. If the average tar content of cigarettes remains constant at the 1986 level, the Health of the Nation targets will not to be met for either males or females. In this model the effects of reducing either the percentage of the population smoking or the number of cigarettes smoked per person are probably underestimated (Section 10.2).

It is assumed that the smoking behaviour of the birth cohort in the 25-34 year age group is an indicator of the life experience of the birth cohort. This model does not allow for differential changes in smoking behaviour among birth cohort, such as patterns of quitting or changing of the type of cigarette smoked. If better information on the patterns of smoking behaviour, were available with particular reference to quitting, the effects on cancer incidence of achieving the Health of the Nation targets for smoking behaviour could be better quantified. The Health of the Nation strategy does not set targets for reduction of the tar content of cigarettes. Since this factor appears to play a vital role in the determination

of the level of lung cancer incidence it is unfortunate that the Health of the Nation strategy sets no targets for this factor. Since 1988 it has not been possible to monitor the trends in this indicator (Section 10.2).

To fully understand the effect of achieving the Health of the Nation targets for smoking on lung cancer incidence information on the changing patterns of smoking are also needed. Since the composition of cigarettes also plays a major role in determining lung cancer incidence there is a need to set targets for the reduction of the average tar content of cigarettes. To assist in the monitoring of the achievement of such a target, it is also necessary for information on the sales adjusted average tar content of cigarettes to be published.

#### **10.4 Implications for Service Needs**

The Report of the Chief Medical Officer's Expert Advisory Group on Cancer has developed a framework for the rationalisation of cancer services within England. This report has emphasised the necessity of cancer services being accessible to the patients, therapy regimens being more standardised, and ensuring an adequate patient through put to maintain specialist expertise. The Association of Cancer Physicians (1994) have recommended a work load of approximately 200 new patients a year for each consultant medical oncologist. The Royal College of



Radiologists (1986) recommend a work load of 350 new patients per consultant radiotherapist. This is far lower than the average workload per consultant in the UK in 1991 of 560 new patients (Royal College of Radiologists 1991). Projections of future cancer incidence, by site, will be necessary if the recommendations of this report are to be put into practice.

The service needs that have been presented in this thesis are based on the assumption that the therapy regimens for lung cancer patients will not change markedly from those used between 1983 and 1987 in Mersey Region. The type of treatment a new cancer patient may receive appears to be dependant on age. The younger the patient, the more likely they are to receive specific treatment. It is felt that this variation with age is related to the stage of the disease at diagnosis, and the ability of the patient to undergo the therapy. In the South Thames East region there is also a decline over age in the percentage of patients who receive specific therapies (Thames Cancer Registry 1991). However, patients in the South Thames East region are almost twice as likely to receive chemotherapy or radiotherapy in Mersey Region.

Approximately the same percentage of patients receive surgery in both regions. While no inferences can be made about these differences in therapy regimens, the implementation of the report of Chief Medical Officer's Expert Advisory Group on Cancer may change treatment

patterns in MRHA.

The variation in projected service needs in MRHA at the beginning of the 21st century is small relative to the overall variation in the projections. The smallest figures are for the models relating the age-standardised rates to information on smoking behaviour. These projections are known to be inaccurate because of the over-estimate of the rate of decrease of lung cancer incidence (Section 9.4). Also, the age-standardised lung cancer incidence rates do not allow the age-specific treatment rates to be used in the calculation of projected service needs. The projections based on the extrapolation of the broad-band age-specific incidence rates are higher than those based on the extrapolation of the parameter estimates from age-period-cohort modelling. This reflects the reduction in lung cancer incidence in the more recent cohort that is not allowed for in the broad-band-age-specific analysis. The inclusion of smoking effects as cohort effects moderates the projected increase in the younger age groups. This would lead to a lowering of requirements for treatment.

This thesis has only considered the treatment needs for lung cancer patients. However, service needs include diagnostic, treatment (both curative and palliative) and other support services. At present just over half lung cancer patients (54.9% of males and 52.4% of females)

have their diagnosis confirmed with either histology or cytology. This percentage has been increasing slowly over the last few years. Less than 10% of diagnoses (8.9% males and 7.5% females) are confirmed with macroscopic evidence, and the rest are diagnosed on clinical evidence alone. The percent of patients having their diagnosis confirmed either microscopically or macroscopically decreases with age. This may reflect the fitness of the patients to undergo such tests. Females are somewhat less likely to have confirmatory test than males. If service needs for diagnosis were to be projected, age specific lung cancer incidence rates are also required.

Any projection of future treatment needs assumes that treatment patterns will remain the same. Although therapy regimens are known to vary throughout the country, there is little or no information on trends in therapy for lung cancer patients in Mersey Region, in recent years. It therefore does not seem profitable to hypothesise about changes to therapy regimens by the year 2000. The data used in this thesis are the most recently available data, and are used to demonstrate a simple method to estimate future service needs in Mersey Region.

Because of the short median survival time for lung cancer service needs can be estimated from incidence. If the survival time were longer both incidence and prevalence would be required.



For many lung cancer patients, treatment is concerned with relieving suffering rather than attempting cure, because of the stage of the disease at presentation. At present the patients who do receive specific treatment, receive one or a combination of surgery, chemotherapy and/or radiotherapy (Williams et al 1993). Approximately 10% of patients receive surgery, 10% receive chemotherapy and 18% receive radiotherapy. Just under 60% of patients receive no specific treatment. For approximately 5% of patients the treatment they receive is not known to the registry. Whether a patient receive a specific therapy and what type of therapy appears to depend on the age of the patient. The younger the patient the more likely he or she is to have therapy. However, this relationship is probably due to the stage of the disease at presentation, and the fitness of the patient to undergo the therapy. Thus if future service needs are to be estimated it is necessary to be able to predict age specific rates and frequencies for lung cancer.

The projection of service needs, based on 1983-87 therapy regimens, imply that for lung cancer alone MRHA would require the services of one full time equivalent consultant in medical oncology and one full time equivalent consultant radiotherapist by the turn of the century.

## 10.5 Conclusions

The Canadian workshop on projections recommended that the projection methods should be as simple as possible, and based on as few assumptions as is necessary (Mclaughlin, Morgan and Mao 1992). The simplest models developed in this thesis relate changes in cancer incidence to changes over calendar period. These models are refined by the inclusion of cohort effects, through age-period-cohort analyses. In order to improve the accuracy of these models, information on smoking behaviour is also included. These models also allow the investigation of the effects of changes in smoking behaviour on lung cancer incidence to be investigated.

Two models perform well for projecting future lung cancer incidence in Mersey Region. These are the models based on age-period-cohort analysis incorporating information on the average tar content of cigarettes and the percentage of the population who smoke or the number of cigarettes smoked per smoker. These models allow the trends in lung cancer incidence over cohorts to be investigated along with the effects of smoking behaviour.

Using either of these models lung cancer incidence in males is projected to continue its current decline and this decrease up to 2011 will meet the targets set by the Health of the Nation. In females, the age-specific lung

cancer incidence rates are falling in the younger age groups, but are still continuing to rise in the older age groups. These changes can be explained by trends over cohorts. The pattern of trends in lung cancer incidence in females is similar to the pattern of trends in males 30 years previously. Unfortunately the projections from these models are that the Health of the Nation target for the reduction of lung cancer in females will not be met.

The models could possibly be improved if information on quitting smoking were more readily available. The trends in the composition of cigarettes are also a major determinant of trends in lung cancer incidence. Targets should be set for the continued reduction of the average tar content of cigarettes. Data on the market share of brands of cigarettes should be published so that the indicator of the average tar content of cigarettes can continue to be monitored.

The treatment a cancer patient receives appears to depend on the age at diagnosis. This may be due to the perceived ability of the patients to withstand treatment. Using data on the pattern of treatments given from 1983-87, the projections indicate that at the turn of the century there would be enough new lung cancer patients in MRHA to justify one full time equivalent consultant medical oncologist and one full time equivalent consultant radiotherapist.



The models proposed in this thesis could be generalised to other registries for lung cancer incidence, using national information on smoking behaviour. For other cancer sites the traditional age-period-cohort modelling (Chapter 8) is always an option, but this thesis has demonstrated a useful methodology for incorporating risk factor data when they are available.

## REFERENCES

Alavanja M.C.R., Brown C.C., Swanson C., Brownson R.C.  
(1993) "Saturated Fat Intake and Lung Cancer Risk Among  
Nonsmoking Women in Missouri" Journal of the National  
Cancer Institute 85:1906-16

Alavanja M.C.R., Brownson R.C., Lubin J.H., Berger E.,  
Chang J., Boice Jr. J.D. (1994) "Residential Radon  
Exposure and Lung Cancer among Nonsmoking Women" Journal  
of the National Cancer Institute. 86:1829-1837

The Alpha-Tocopherol, Beta Carotene cancer prevention  
study group (1994) "The effect of Vitamin and beta  
carotene on the incidence of lung cancer and other  
cancers in male smokers" The New England Journal of  
Medicine; 330:1029-35

Anonymous "Brand Shares of the UK Cigarette Market"  
Tobacco July 1973:26 July 1974:39 July 1975:19 July  
1976:40-1 August 1977:18-19 August 1979:20-1 August  
1980:28-9 August 1981:24-5 August 1982:18-9 September  
1983:20-1 September 1984:7 September 1985:4-5 August  
1986:7-9 August 1987:15-17 August 1988:7-9 August  
1989:10-11

Association of Cancer Physicians (1994) Review of the  
Pattern of Cancer Services in England and Wales.



Baker R.J., Nelder J.A. (1978) "The GLIM System", Release 3, Oxford, Numerical Algorithms Group

Barrett J.C. (1978) "The Redundant Factor Method and Bladder Cancer Mortality" *Journal of Epidemiology and Community Health* 32:314-316

Breslow N.E. (1984) "Extra-Poisson Variation in Log-Linear Models" *Applied Statistics* 33,1:38-44

Brown C.C., Kessler L.G. (1988). "Projections of Lung Cancer Mortality in the United States: 1985-2025" *Journal of the National Cancer Institute*. 80:43-51.

Byers T., Vena J., Mettlin C., Swanson M., Graham S. (1994) "Dietary Vitamin A and lung cancer risk: an analysis by histologic subtypes" *American Journal of Epidemiology*. 120:769-76

Case R.M. (1956) "Cohort Analysis of Mortality Rates as an Historical or Narrative Technique" *British Journal of Preventative and Social Medicine* 10:159-71

Chaffey C.M., Bowie C. (1994) "Radon and health - an update" *Journal of Public Health Medicine* 16:465-70

Clayton D. Schifflers E. (1987a). "Models for Temporal Variations in Cancer Rates 1. Age-Period and Age-Cohort Models". *Statistics in Medicine*. 6:449-467.

Clayton D. Schifflers E. (1987b) "Models for Temporal Variation in Cancer Rates. II Age-Period - Cohort Models". *Statistics in Medicine*. 6:468-481.

Clayton D., Hills M. (1993) "Statistical Models in Epidemiology" Oxford University Press, Oxford

Clemmesen J. (1965) "Statistical Studies in the Aetiology of Malignant Neoplasms I:Review and Results, II:Basic Tables, Denmark 1943-1957" *Acta Pathology and microbiology Scandinavia Supplement* 174,1965

Clemmesen J., Busk T. (1947) "On the Apparent Increase in the Incidence of Lung Cancer in Denmark, 1931-45" *British Journal of Cancer* 1:253-259

Day N.E. Charney B. (1982). "Time Trends, Cohort Effects and Ageing as Influence on Cancer Incidence". In Magnus K. Ed; "Trends in Cancer Incidence" New York, Hemisphere. pp 51-65.

Dean G., Lee P.N., Todd G.F., Wicken A.J. (1978) "Report on the second retrospective Mortality study in North Eastern England. Part II Changes in Lung Cancer and Bronchitis Mortality and in Other Relevant Factors Occurring in Areas of North-east England 1963-1972" (Research Paper 14) London, Tobacco Research Council

Department of Health and Social Security (1980) "Inequalities in Health. Report of a Research Working Group (Black Report)" London

Department of Health (1993). "The Health of the Nation, a Consultative Document for Health in England". HMSO, London.

Department of Health (1994) "Expert Advisory Group on Cancer to the Chief Medical Officers of England and Wales. Consultative Document: a Framework for Commissioning Cancer Services" HMSO, London.

De Vos Irvine H., Lamont D.W., Hole D.J., Gillis C.R., (1993) "Asbestos and Lung Cancer in Glasgow and the West of Scotland" British Medical Journal 306:1503-6

Doll R., (1953) "Mortality from lung cancer among non-smokers". Preliminary Report" British Journal of Cancer 7:3039-312



Doll R. (1971) "Age Distributions of Cancer: implications for models of carcinogenesis" Journal of the Royal Statistical Society A 134:133-166

Doll R., Hill A.B., (1950) "Smoking and carcinoma of the lung. Preliminary Report" British Medical Journal ii:739-748

Doll R., Hill A.B., (1964) "Mortality in Relation to Smoking: Ten years' observation of British Doctors" British Medical Journal 1:1399-1410

Doll R., Peto R., (1976) "Mortality in Relation to Smoking: 20 years' observation of male British Doctors" British Medical Journal ii:1525-1536

Doll R., Gray R., Hafner B., Peto R., (1980) "Mortality in Relation to Smoking: 22 years' observation of female British Doctors" British Medical Journal 280:967-971

Draper N.R. Smith H (1981) "Applied Regression Analysis" Wiley, New York

Fleiss J.L. (1981). "Statistical Methods for Rates and Proportions", Wiley, New York.

Fontham E.T.H., (1990) "Protective Dietary Habits and Lung Cancer" International Journal of Epidemiology 19:S32-S42

Forman M.F., Yao S.X., Graubard B.I., Qiao Y.L., McAdams M., Mao B.L., Taylor P.R. (1992) "The effect of dietary intake of fruits and vegetables on the odds ratio of lung cancer among Yunnan Tin Miners" International Journal of Epidemiology 21:437-441

Friberg L., Cedarlof R. (1978) "Late effects of air pollution with special reference to lung cancer" Environmental Health Perspective 22:45-66

Garfinkel L. (1980) "Cancer Mortality in NON-smokers: Prospective Study by the American Cancer Society" Journal of the National Cancer Institute 65:1169-73

Garfinkel L., Stellman S.D. (1988) "Smoking and Lung Cancer in Women: findings in a prospective study" Cancer Research 48:6951-6955

Greaves J.P., Hollingsworth D.F. (1966) "Trends in Food Consumption in the United Kingdom" World Review of Nutrition and Dietetics. 6:34-89

Greenwald P., Sondik E., Eds (1986) "Cancer Control Objectives for the Nation: 1985-2000" (National Cancer Institute Monograph No.2), Washington DC, US Department of Health and Human Services.

Hakama M., Pukkala E. (1984). "The Projection of Chronic Diseases using Data on Risk Factors and Risk Factor Intervention: The Case of Cancer". World Health Statistics Quarterly. 3:318-327.

Hakama M., Hakulinen T., Laara E. (1986). "Predicting Cancer Incidence and Prevalence". In Health Projections in Europe. Methods and Applications. World Health Organisation, Regional Office for Europe, Copenhagen. pp 25-39.

Hakama M., Beral V., Cullen J., Parkin M. (1989) "UICC Workshop on Evaluating Interventions to Reduce Cancer Risk" International Journal of Cancer 43:967-969

Hakulinen T., Pukkala E. (1981). "Future Incidence of Lung Cancer: Forecasts Based on Hypothetical Changes in the Smoking Habits of Males". International Journal of Epidemiology ,10:233-240



Hakulinen T., Laara E., Pukkala E. (1985) "Prediction of Lung Cancer Incidence in Finland: Appraisal of different approaches". In: Proceedings of the 5th World Conference on Smoking and Health, Winnipeg, Canada, 1983. Vol. 1. Ed: Forbes W.F., Frecker R.C., and Nostbakken D., Canadian Council on Smoking and Health, Ottawa. pp 707-720.

Hakulinen T., Teppo L., Saxen E. (1986). "Do the Predictions for Cancer Incidence Come True? Experience from Finland". Cancer. 57:2454-2458

Hakulinen T., Pukkala E., Kenward M., Teppo L., Puska P., Tuomilehto J., Kuulasmaa K. (1990) "Changes in Cancer Incidence in North Karelia, an Area with a Comprehensive Preventative Cardiovascular Programme" In "Evaluating Effectiveness of Primary Prevention of Cancer" Eds: Hakama M., Beral V., Cullen J.W., Parkin D.M., IARC Scientific Publications No. 103:133-148

Hakulinen T., Dyba T. (1994) "Precision of Incidence Predictions based on Poisson Distributed Observations" Statistics in Medicine 13:1513-1523

Hammond E.C. (1966) "Smoking in Relation to the Death Rates of One Million Men and women" National Cancer Institute Monograph 19:127-204

Hammond E.C., Selikoff I.J., Seidman H. (1979) "Asbestos Exposure, Cigarette Smoking and Death Rates" Annals of the New York Academy of Science 330:473-490

Harris J (1983) "Cigarette Smoking among Successive Birth Cohorts of men and women in the United States during 1900-1980" Journal of the National Cancer Institute 71:473-479

Harrison P.J., Stevens C.F. (1976) "Bayesian Forecasting" Journal of the Royal Statistical Society 38:205-247

Hayes R.B., Vineis P. (1989) "Time Dependency in Human Cancer" Tumori 75:189-195

Health Departments of the United Kingdoms "Tar and Nicotine Yields of Cigarettes" Leaflets issued in April 1973, February 1974, September 1974, May 1975, January 1976, August 1976, March 1977, October 1977, July 1978, January 1979, October 1979, May 1980, March 1981, October 1981, April 1982.

Hennekens C.H., Buring J.E., Manson J.E., Stampfer M., Rosner B., Cook N.R., Belanger C., LaMotte F., Gaziano J.M., Ridker P.M., Willett W., Peto R. (1996) "Lack of effect of long-term supplementation with beta carotene on the incidence of malignant neoplasms and cardiovascular disease" The New England Journal of Medicine; 334:1145-49

Holford T.R. (1983). "The estimation of age, period and cohort effects for vital rates", *Biometrics*. 39:311-324.

Holford T.R. (1985) "An Alternative Approach to Statistical Age-Period-Cohort Analysis" *Journal of Chronic Disease* 38,10:831-840

Holford T.R., Zhang Z., McKay L. (1994) "Estimating Age, Period and Cohort Effects using the Multistage Model for Cancer" *Statistics in Medicine* 13:23-41

Hornung R.W., Meinhardt T.J. (1987) Quantitative Risk Assessment of lung cancer in US Uranium Miners. *Health Physician* 52:417-430

Howe G.R., Nair R.C., Newcombe H.B., Miller A.B., Abbatt J.D. (1986) "Lung Cancer mortality (1950-80) in relation to radon daughter exposure in a cohort of workers at the Eldorado Beaverlodge uranium mine" *Journal of the National Cancer Institute*. 77:357-62

IARC (1985) IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans Vol 38 Tobacco Smoking, Lyon



James I.R., Segal M.R. (1982) "On a Method of Mortality Analysis Incorporating Age-Year Interaction, with Application to Prostate Cancer Mortality" *Biometrics* 38:433-43

Juel K. (1983) "Demographic Factors and Cancer Mortality. A Mathematical Model for Cancer Mortality in Denmark 1943-1978" *International Journal of Epidemiology* 12:419-425

Kaldor J., Harris J.A., Glazer E., Glaser S., Neutra R., Mayberry R., Nelson V., Robinson L., Reeds D. (1984) "Statistical associations between cancer incidence and major-cause mortality and estimated residential exposure to air emissions from petroleum and chemical plants" *Environmental Health Perspective* 54:319-332

Kupper L.L., Janis J.M., Karmous A., Greenberg B.G. (1984) "Age-Period-Cohort Analysis: a Review and Critique" *Journal of Chronic Diseases* 38,10:811-830

Kusiak R.A., Ritchie A.C., Muller J., Springer J. (1995) "Mortality from Lung Cancer in Ontario Uranium Miners" *British Journal of Industrial Medicine* 50:920-28

La Vecchia C., Levi F., Decarli A., Wietlisbach V., Negri E., Gutzwiller F. (1988) "Trends in Smoking and Lung Cancer Mortality in Switzerland" *Preventive Medicine* 17:712-724

Lawther P.J., Waller R.E. (1978) "Trends in Urban Air Pollution in the United Kingdom in relation to Lung Cancer Mortality" *Environmental Health Perspectives* 22:71-73

Lees R.E.M., Steele R., Roberts J.H., (1987) "A case-control study of lung cancer relative to domestic radon exposure" *International Journal of Epidemiology* 16:1-12

Le Marchand L., Hankin J.H., Bach F., Kolonel L.N., Wilkens L.R., Stacewicz-Sapuntzakis M., Bowen P.E., Beecher G.R., Laudon F., Baque P., Daniel R., Seruvatu L., Henderson B.E. (1995) "An ecological study of diet and lung cancer in the South Pacific" *International Journal of Cancer* 63:18-23

Letourneau E.G., Krewski D., Choi N.W., Goddard M.J., McGregor R.G., Zielinski J.M., Du J. (1994) "Case-Control Study of Residential Radon and Lung Cancer in Winnipeg, Manitoba, Canada. 140:310-22

Levin D.L., Gail M.H., Kessler L.G., Eddy D.M. (1986) "A Model for Projecting Cancer Incidence and Mortality in the Presence of Prevention, Screening and Treatment Programs" National Cancer Institute Monographs 2:83-93

Levin M.L., Haenszel W., Carrol B.E., Gerhardt P.R., Handy V.H., Ingraham S.C. (1960) "Cancer incidence in Urban and rural areas of New York State" Journal of the National Cancer Institute 24:1243-1257

Lloyd O.L.L. Williams F.L.R., Gailey F.A.Y. (1985) "Is the Armedale epidemic over? Air pollution and mortality from lung cancer and other diseases, 1961-1982" British Journal of Industrial Medicine 42:815-823

Lubin J.H., Blot W.J., Berrino F., Lamant F., Gilles C.R., Kunze M., Schmahl D., Visco G. (1984a) "Modifying risk of developing lung cancer by changing habits of cigarettes smoking" British Medical Journal 288:1953-1956

Lubin J.H., Blot W.J., Berrino F., Flamant R., Gilles C.R., Kunze M., Schmahl D., Visco G. (1984b) "Patterns of lung cancer risk according to type of cigarette smoked" International Journal of Cancer 33:569-576



Lubin J.H., Boice Jr. J.D., Edling C., Hornung R.W., Howe G.R., Kunz E., Kusiak R.A., Morrison H.I., Radford E.P., Samet J.M., Tirmarche M., Woodward A., Yao S.X., Pierce D.A. (1995) "Lung Cancer Radon-Exposed Miners and Estimation of Risk from Indoor Exposure" Journal of the National Cancer Institute 87:817-27

McCullagh M., Nelder J.A. (1989) "Generalised Linear Models" 2nd Edition, Chapman Hall, London

McLaughlin J.R., Morgan P., Mao Y. (1992) "Executive Summary: First Canadian Cancer Projection Workshop" Chronic Diseases in Canada 13(3):47-51

Mancuso T.F., MacFarlane E.M., Porterfield J.D. (1955) "Distribution of cancer mortality in Ohio" American Journal of Public Health 45:58-70

Marrett L.D. Clarke E.A. Hatcher J. Weir H.F. (1986). "Epidemiological Research using the Ontario Cancer Registry" Canadian Journal of Public Health. 77(1):79-85.

Mayne S.T., Janerich D.T., Greenwald P., Chorost S., Tucci C., Zaman M.B., Melamed M.R., Kiely M., McKneally M.F. (1994) "Dietary Beta Carotene and Lung Cancer Risk in US Nonsmokers" Journal of the National Cancer Institute 86:33-38

Ministry of Health (1954) "Memorandum on Cancer Registration - National Health Services" Ministry of Health, Saville Row, London

Morrison H.I., Semenciw R.M., Mao Y., Wigle D.T. (1998) "Cancer Mortality among a group of Fluorspar miners exposed to radon progeny" American Journal of Epidemiology. 128:1266-75

Morrison H.I., MacNeill I.B., Miller D., Levy I., Xie l., Mao Y. (1995) "The impending Canadian Prostate Cancer Epidemic" Canadian Journal of Public Health 86(4):274-278

Muir C., Waterhouse J., Mack T., Powell J., Whelan S. eds (1987) "Cancer Incidence in five Continents Vol V" (IARC Publications No 88) Lyon IARC

Multiple Risk Factor Intervention Trial Research Group (1982) "Multiple Risk Factor Intervention Trial. Risk Factor Changes and Mortality Results" Journal of the American Medical Association 248(12):1465-1477

National Centre for Health Statistics (1984) "Health, United States" Washington, DC: US Govt Print Off, 1984 [DHHS publication No. (PHS) 84-1232]

Negri E., La Vecchia C., Levi F., Randriamiharisoa A., Decarli A., Boyle P. (1990) "The Application of Age, Period and Cohort models to Predict Swiss Cancer Mortality" *Journal of Cancer Research and Clinical Oncology* 116:207-214

Omenn G.S., Goodman G.E., Thornquist M.D., Balmes J., Cullen M.R., Glass A., Keogh J.P., Meyskens F.L., Valanis B., Williams J.H., Barnhart S., Hammar S. (1996) "Effects of a combination of beta carotene and vitamin A on lung cancer and cardiovascular disease" *The New England Journal of Medicine*; 334:1150-55

Ontario Cancer Treatment and Research Foundation (1986) "Estimates of the number of new cancer cases in Ontario in 1991, 1996 and 2001, together with estimates of the number of radiotherapy machines required" Internal Report

OPCS (1954a) "Census 1951, County Report Lancashire" London, HMSO, Tables 21 and 22

OPCS (1954b) "Census 1951, County Report Cheshire" London, HMSO, Tables 21 and 22

OPCS (1964a) "Census 1961, County Report Lancashire" London, HMSO, Table 6

OPCS (1964b) "Census 1961, County Report Cheshire"



London, HMSO, Table 6

OPCS (1966a) "Census 1961, Occupation, Industry and Socio Economic Groups, Lancashire" London HMSO Tables 1 and 3

OPCS (1966b) "Census 1961, Occupation, Industry and Socio Economic Groups, Cheshire" London HMSO Tables 1 and 3

OPCS (1976) "Census 1971, Economic Activity, Sub Regional Tables (10% sample)" London HMSO Table 1

OPCS (1984a) "Census 1981, Economic Activity, Lancashire" London HMSO Table 2

OPCS (1984b) "Census 1981, Economic Activity, Cheshire" London HMSO Table 2

OPCS (1995) "1989- based Sub-national Population Projection", Series PP3 no. 8.

Osmond C. (1985). "Using Age-Period-Cohort models to estimate Future Mortality rates" International Journal of Epidemiology. 14:124-129.

Osmond C., Gardner M.J. (1982). "Age, period and cohort models, applied to cancer mortality rates" Statistics in Medicine. 1:245-259.

Osmond C., Gardner M.J., Acheson E.D. (1983) Trends in

Cancer Mortality 1951-1980. Analysis by Period of Birth and Death" Her Majesty's Stationery Office, London

Pastorino U., Berrino F., Gervasio A., Persenti V., Riboli E., Crosignani P (1984) "Proportion of lung cancers due to occupational exposures" International Journal of Cancer 33:231-237

Pershagen G. (1985) "Lung Cancer Mortality among men living near an arsenic emitting smelter" American Journal of Epidemiology 122:684-689

Pershagen G., Akerblom G., Axelson O., Clavensjo B., Damber L., Desai G., Enflo A., Lagarde F., Mellander H., Svartengren M., Swedjemark G.A. (1994) "Residential Radon Exposure and Lung Cancer in Sweden" New England Journal of Medicine. 330:159-64

Peto R., Lopez A.D., Boreham J., Thun M., Heath C.Jr. (1992) "Mortality from Tobacco in developed Countries: indirect estimation from national vital statistics" The Lancet 339:1268-1267

Pukkala E., Teppo L., Hakulinen T., Rimpela M. (1983) "Occupation and Smoking as Determinants of Lung Cancer" International Journal of Epidemiology 12:290-6

Puska P (1973) "The North Karelia Project: an attempt at

community prevention of cardiovascular disease" WHO  
Chronicle 27:55-58

Radford E.P., Renard K.G.S.C. (1984) "Lung Cancer in  
Swedish iron miners exposed to low doses of radon  
daughters" New England Journal of Medicine. 310:1485-1494

The Registrar General (1963) "Statistical Review of  
England and Wales, for the Year 1972 Part II Tables,  
Population", London, HMSO, Table A4.

The Registrar General (1964) "Statistical Review of  
England and Wales, for the Year 1972 Part II Tables,  
Population", London, HMSO, Table A4.



The Registrar General (1965) "Statistical Review of England and Wales, for the Year 1973 Part II Tables, Population", London, HMSO, Table A4.

The Registrar General (1967) "Statistical Review of England and Wales, for the Year 1965 Part II Tables, Population", London, HMSO, Table A4.

The Registrar General (1968) "Statistical Review of England and Wales, for the Year 1966 Part II Tables, Population", London, HMSO, Table A4.

The Registrar General (1969) "Statistical Review of England and Wales, for the Year 1967 Part II Tables, Population", London, HMSO, Table A4.

The Registrar General (1970) "Statistical Review of England and Wales, for the Year 1968 Part II Tables, Population", London, HMSO, Table A4.

The Registrar General (1971) "Statistical Review of England and Wales, for the Year 1969 Part II Tables, Population", London, HMSO, Table A4.

The Registrar General (1972) "Statistical Review of England and Wales, for the Year 1970 Part II Tables, Population", London, HMSO, Table A4.

The Registrar General (1973) "Statistical Review of England and Wales, for the Year 1971 Part II Tables, Population", London, HMSO, Table A4.

Reissigova J., Luostarinen T., Hakulinen T., Kubik A.  
"Statistical Modelling and Prediction of Lung Cancer Mortality in the Czech and Slovak Republics, 1960-1999"  
International Journal of Epidemiology 23:665-672

Rimpela A.H., Pukkala E.I. (1987) "Cancers of Affluence: Positive Social Class Gradient and Rising Incidence Trend in Some Cancer Forms" Social Science and Medicine 24:601-606

Risch H.A., Howe G.R., Jain M., Burch J.D., Holowaty E.J., Miller A.B. (1993), "Are Female Smokers at higher risk for Lung Cancer than Male Smokers". The American Journal of Epidemiology 138:281-293.

Rodgers W.L. (1982) "Estimable Functions of Age, Period and Cohort Effects" American Sociological Review 47:774-787

Rose G., Hamilton P., Colwell L., Shapely M. (1982) "A randomised control trial of anti-smoking advice: 10 year results" Journal of Epidemiology and Community Health 36:102-108

Roush G.C., Schymira M.S., Holford T.R., White C.,  
Flannery J.T., "Time Period compared to Birth Cohort in  
Connecticut Incidence Rates for Twenty Five Malignant  
Neoplasms" Journal of the National Cancer Institute  
74(4):779-88

Royal College of Physicians (1970) "Air Pollution and  
Health" London, Putnam Medical

Royal College of Radiologists (1986) "Working Group on  
Workload Assessment and Manpower"

Royal College of Radiologists (1991) "Medical Manpower  
and Workload in Clinical Oncology in the UK"

Schoenberg J.B., Klotz J.B., Wilcox G.P., Nicholls G.P.,  
Gil del Real M.T., Stemhagen A., Mason T.J. (1990) "Case-  
control study of residential radon and lung cancer among  
New Jersey women" Cancer Research. 50:6250-54

Smith G.H., Williams F.L., LLOYD O.L. (1987) "Respiratory  
Cancer and air pollution from iron foundries in a  
Scottish Town: an epidemiological and environmental  
study" British Journal of Industrial Medicine 44:795-802

Steinmetz K.A., Potter J.D., Folsom A.R. (1993)  
"Vegetables, Fruit and Lung Cancer in the Iowa Women's  
Health Study" Cancer Research 53:536-43



Stevens R.G. Moolgavkar S.H. (1979) "Estimation of Relative Risk from Vital Data: Smoking and Cancers of the Lung and Bladder" Journal of the National Cancer Institute 63:1351-1357

Stevens R.G. Moolgavkar S.H. (1984) "A cohort analysis lung cancer and smoking and Cancers in British Males" American Journal of Epidemiology 119:624-641

Stocks P. (1947) "Studies on Medical and Population Subjects, No. 1. Regional and Local Differences in Cancer Death Rates" HMSO London

Stocks P., Campbell J.M (1955) "Lung cancer death rates among non smokers and pipe and cigarette smokers: evaluation in relation to air pollution by benzpyrene and other substances" British Medical Journal ii:923-929

Teppo L. (1984). "Lung Cancer in Scandinavia: Time Trends and Smoking Habits" In: Mizell M., Correa P., eds. Lung Cancer: Causes and Prevention. Deerfields Beach: Chemie International. pp 21-31

Teppo L., Hakulinen T., Saxen E. (1974) "The Prediction of Cancer in Finland for the year 1980 by Means of Cancer Registration Material" Annals of Clinical Research 6:122-125

Thames Cancer Registry (1991) "Cancer in East and West

Kent HAs, 1991: Incidence, prevalence, survival and treatment patterns for patients resident in the Health Authorities of East and West Kent" A report for Dr A.E. Limentani, Director of Public Health, East Kent Health Authority

Thorn M., Sparen P., Bergstrom R., Adami H.-O. (1992) "Trends in Mortality Rates from Malignant Melanoma in Sweden 1953-1987 and forecasts up to 2007" British Journal of Cancer 66:563-567

Tomatis L. Ed. (1990). "Cancer Causes, Occurrence and Control". WHO, IARC Scientific Publications No. 100.

Townsend J.L. (1978) "Smoking and Lung Cancer: A cohort Data Study of Men and Women in England and Wales 1935-70" Journal of the Royal Statistical Society, A 141(1):95-107

Vioque J., Bolumar F. (1987) "Trends in Mortality from Lung Cancer in Spain, 1951-1980" Journal of Epidemiology and Community Health 41:74-78

Wald N., Doll R., Copeland G. (1981) "Trends in Tar, Nicotine, and Carbon Monoxide yields of UK cigarettes manufactured since 1934" British Medical Journal 282:763-765

Wald N., Nicolaidis-Bouman A. Ed. (1991). "UK Smoking

Statistics" Wolfson Institute of Preventive Medicine and Oxford University Press, Oxford.

Waterhouse J., Muir C., Correa P., Powell J. Ed. (1976) "Cancer Incidence in Five Continents, Volume III" IARC Scientific Publications No 15. Lyon:International Agency for Research on Cancer

Whittemore A.S., McMillan A. (1983) Lung Cancer Mortality among US uranium miners: a reappraisal. Journal of the National Cancer Institute 71:489-499

WHO "International Classification of Diseases for Oncology" WHO, Geneva 1976

WHO Technical Report Series, No. 632. (1979) "Cancer Statistics" report of a WHO/IARC Expert Committee.

Williams E.M.I., Youngson J.H., Ashby D., Donnelly R.J. Eds (1993) "Lung Cancer Bulletin" Mersey Regional Cancer Registry.

Wynder E.L., Graham E.A. (1950) "Tobacco smoking as a possible etiological factor in bronchogenic carcinoma. A study of six hundred and eighty four proven cases" Journal of the American Medical Association 143:329-336



## APPENDICES

These appendices are numbered so that the roman numeral relates directly to the Chapter number. There are no appendices for Chapters 1 or 2

APPENDIX III(1)

Fifth Digit Morphology Codes

5th Digit	Behaviour
/0	Benign
/1	Uncertain whether benign or malignant Borderline malignancy
/2	Carcinoma-in-situ Intraepithelial Non-infiltrating Non-invasive
/3	Malignant, primary site
/6	Malignant, Metastatic site Secondary site
/9	Malignant, uncertain whether primary or metastatic site
/A	Primary known, report applies to secondary site only

APPENDIX III(2a)

Residence Codes Lancashire

Local Authority	GRO	OPCS
Bootle C.B.	8710	64JA
Liverpool C.B.	8700	64C*
St Helens C.B.	8800	64GA
Warrington C.B.	8811	63PA
Southport C.B.	8810	64JC
Crosby M.B.	8730	64JE
Widnes M.B.	8831	63JE
Formby U.D.	8860	64JF
Golbourne U.D.	8840	63PC
Haydock U.D.	8861	64GE
Huyton with Roby U.D.	8731	64AA
Kirkby U.D.	8830	64AB
Litherland U.D.	8740	64JG
Ormskirk U.D.	8842	68PA
Newton-le-Willows U.D.	8841	64GF
Prescot U.D.	8862	64AC
Rainford U.D.	8863	64GG
Skelmersdale and Holland U.D.	8869	68PC
Warrington R.D.	8880	63PP, 63PR
West Lancashire R.D.	8881	64JL, 64AL
Whiston R.D.	8882	63JP, 63PT, 64GL, 64AN

APPENDIX III (2a)

Residence Codes Cheshire

Local Authority	GRO	OPCS
Birkenhead C.B.	8500	64LA, 64LB
Chester C.B.	8610	63AA
Wallasey C.B.	8501	64LC
Bebington M.B.	8530	64LD
Congleton M.B.	7962	63CC
Crewe M.B.	7930	63EA
Ellesmere Port M.B.	8540	63GA
Macclesfield M.B.	7940	63LG
Alderley Edge U.D.	7860	63LA
Alsager U.D.	7960	63CA
Ashton U.D.	8240	64GC
Billinge U.D.	8265	64GD
Bollington U.D.	7961	63LC
Buckley U.D.	9470	71AA
Connah's Quay U.D.	9471	71AC
Hoylake U.D.	8541	64LH, 64LJ
Knutsford U.D.	7963	63LE
Lymm U.D.	8660	63PE
Middlewich U.D.	7965	63CE
Nantwich U.D.	7966	63EC
Neston U.D.	8560	63GC
Northwich U.D.	7967	63NA
Runcorn U.D.	8640	63JA, 63JC
Sandbach U.D.	7968	63CG
Winsford U.D.	7969	63NC
Wirral U.D.	8542	64LK
Bucklow R.D.	7980	63LL
Chester R.D.	8680	63AL
Congleton R.D.	7981	63CN, 63CL
Disley R.D.	7880	63LN
Hawarden R.D.	9490	71AL, 71LL
Macclesfield R.D.	7982	63LP
Nantwich R.D.	7983	63EL
Northwich R.D.	7985	63NL
Runcorn R.D.	8682	63JL, 63JN, 63NN, 63NP, 63PL, 63PN
Tarvin R.D.	8683	63AN



## APPENDIX III (2b)

## POSTAL CODES

Area	Postal Code
Wirral	L41 - L49, L60 -L66
Cheshire	CH1 - CH4
Liverpool	L1 - L38
Southport	PR8, PR9
Whiston, Runcorn, St Helens, Knutsford, etc	WA1 - WA16
Crewe, Congleton	CW1 - CW12
Alsager	ST7

APPENDIX III (3a)

Treatments Received by Male Lung Cancer Patients in Mersey Region 1983-87

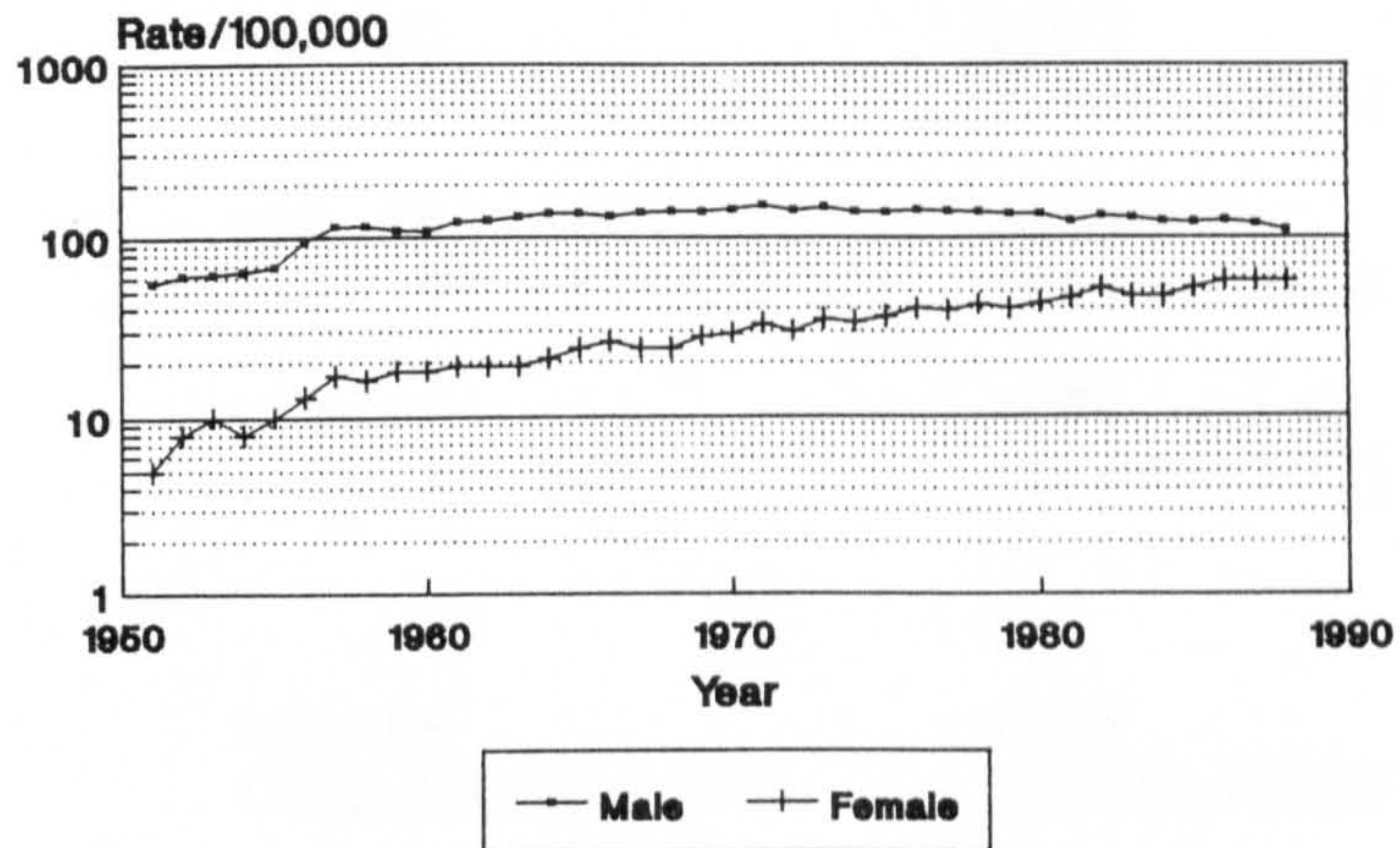
Age (yrs)	Type of Treatment						Total
	Non-Specific	Surgery	Chemo-therapy	Radio-therapy	Not Known		
<55	No.	115	139	169	42		610
	%	(18.9)	(22.8)	(27.7)	(6.9)		
55-64	No.	323	242	480	143		1951
	%	(16.6)	(12.4)	(24.6)	(7.3)		
65-74	No.	281	219	552	201		2910
	%	(9.7)	(7.5)	(20.0)	(6.9)		
75+	No.	31	74	160	83		2019
	%	(1.5)	(3.7)	(7.9)	(4.1)		
Total	No.	750	674	1361	469		7490
	%	(10.0)	(9.0)	(18.2)	(6.2)		

APPENDIX III (3b)

Treatments Received by Female Lung Cancer Patients in Mersey Region 1983-87

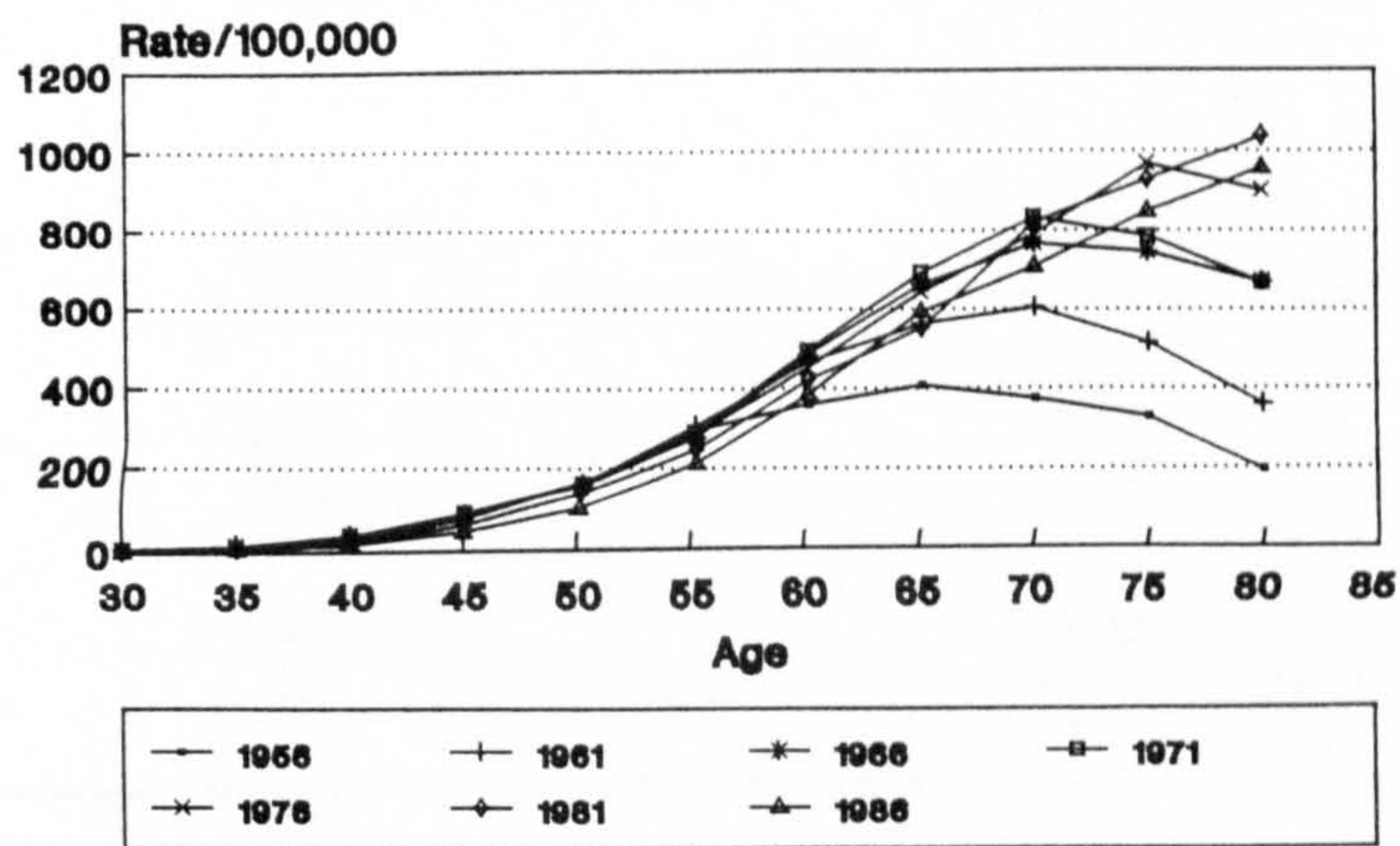
Age (yrs)	Type of Treatment						Total
	Non-Specific	Surgery	Chemo-therapy	Radio-therapy	Not Known		
<55	No.	67	71	82	24	311	
	%	(21.5)	(22.8)	(26.4)	(7.7)		
55-64	No.	131	151	227	53	944	
	%	(13.9)	(16.0)	(24.0)	(5.6)		
65-74	No.	98	109	228	66	1142	
	%	(8.6)	(9.5)	(20.0)	(5.8)		
75+	No.	7	28	52	30	817	
	%	(0.9)	(3.4)	(6.4)	(3.8)		
Total	No.	303	359	589	173	3214	
	%	(9.4)	(11.2)	(18.3)	(5.4)		

**Appendix IV(1) Age-Standardised Lung Cancer Incidence Rates, by Sex**



Standardised to 1981 MRHA Population  
Mersey Region Cancer Registry

**Appendix IV(2) Age-Specific Lung Cancer Incidence Rates by Quinquennia, Male**



Quinquennia centred on census years  
Mersey Regional Cancer Registry



APPENDIX IV(3a)

Projected Male Lung Cancer Incidence. Linear projections of Log Age Standardised Rates against Year (Population for the year 2000 as Standard)

Data Years	Projected Rate/100,000 (C.I.)	Projected Number of Cases (C.I.)
<b>1990</b>		
1964-1988	144.60 (128.30, 162.98)	1687 (1497, 1901)
1969-1988	138.24 (126.02, 151.65)	1613 (1470, 1769)
1974-1988	134.56 (124.10, 145.90)	1570 (1448, 1702)
<b>2000</b>		
1964-1988	137.96 (120.72, 157.67)	1599 (1399, 1828)
1969-1988	125.59 (112.57, 140.11)	1456 (1305, 1624)
1974-1988	117.92 (105.99, 131.19)	1367 (1229, 1521)
<b>2010</b>		
1964-1988	131.63 (113.04, 153.28)	1582 (1358, 1842)
1969-1988	113.98 (99.92, 130.02)	1370 (1201, 1562)
1974-1988	103.34 (89.82, 118.89)	1242 (1079, 1429)

APPENDIX IV(3b)

Projected Female Lung Cancer Incidence. Linear projections of Log Age-Standardised Rates against Year (Population for the year 2000 as Standard)

Data Years	Projected Rate/100,000 (C.I.)	Projected Number of Cases (C.I.)
<b>1990</b>		
1964-1988	66.62 (57.93, 72.29)	800 (716, 894)
1969-1988	65.04 (58.58, 72.21)	804 (724, 893)
1974-1988	64.72 (59.02, 75.20)	824 (730, 930)
<b>2000</b>		
1964-1988	98.49 (86.04, 112.75)	1197 (1046, 1371)
1969-1988	93.50 (82.64, 105.81)	1137 (1004, 1286)
1974-1988	92.39 (79.84, 106.91)	1123 (971, 1300)
<b>2010</b>		
1964-1988	145.62 (124.82, 169.89)	1802 (1545, 2103)
1969-1988	134.4 (115.78, 156.07)	1664 (1433, 1931)
1974-1988	132.03 (108.97, 159.95)	1634 (1349, 1979)

APPENDIX IV(4a)

Projected Male Lung Cancer Incidence in 1990  
Linear Projections of Broad-Band Age-Specific Rates  
against Year

Data Years	Age Group	Projected Rate /100,000 (C.I.)	Projected Number of Cases (C.I.)
1964-1988	30-44	7 (3,11)	18 (9,26)
	45-64	179 (152,205)	464 (395,533)
	65+	753 (704,803)	1080 (1009,1151)
	Total		1561 (1462,1661)
1969-1988	30-44	6 (3,10)	16 (8,26)
	45-64	168 (151,185)	436 (391,481)
	65+	732 (681,782)	1049 (977,1121)
	Total		1501 (1416,1586)
1974-1988	30-44	7 (4,11)	18 (9,28)
	45-64	163 (146,181)	424 (378,470)
	65+	723 (672,775)	1037 (963,1111)
	Total		1479 (1392,1567)

APPENDIX IV (4a continued)

Projected Male Lung Cancer Incidence in 2000  
 Linear Projections of Broad-Band Age-Specific Rates  
 against Year

Data Years	Age Group	Projected Rate /100,000 (C.I.)	Projected Number of Cases (C.I.)
1964-1988	30-44	5 (2,8)	12 (4,20)
	45-64	154 (137,171)	432 (384,480)
	65+	760 (705,816)	1158 (1073,1243)
	Total		1602 (1504,1700)
1969-1988	30-44	4 (1,7)	10 (3,18)
	45-64	135 (118,152)	379 (332,427)
	65+	716 (657,774)	1090 (1001,1178)
	Total		1479 (1379,1580)
1974-1988	30-44	5 (2,10)	14 (4,24)
	45-64	127 (108,145)	356 (305,408)
	65+	695 (630,761)	1059 (959,1159)
	Total		1429 (1328,1542)



APPENDIX IV (4a continued)  
 Projected Male Lung Cancer Incidence in 2010  
 Linear Projections of Broad Band Age Specific Rates  
 against Year

Data Years	Age Group	Projected Rate /100,000 (C.I.)	Projected Number of Cases (C.I.)
1964-1988	30-44	3 (1,6)	8 (2,14)
	45-64	132 (115,148)	427 (374,481)
	65+	768 (704,832)	1317 (1207,1427)
	Total		1752 (1630,1874)
1969-1988	30-44	3 (0,5)	6 (0,12)
	45-64	108 (92,124)	351 (298,403)
	65+	699 (630,768)	1199 (1081,1317)
	Total		1556 (1426,1685)
1974-1988	30-44	4 (0,8)	10 (0,19)
	45-64	98 (84,113)	318 (271,365)
	65+	669 (587,752)	1148 (1006,1290)
	Total		1476 (1321,1630)

APPENDIX IV (4b)

Projected Female Lung Cancer Incidence in 1990  
Linear Projections of Broad Band Age Specific Rates  
against Year

Data Years	Age Group	Projected Rate /100,000 (C.I.)	Projected Number of Cases (C.I.)
1964-1988	30-44	4 (1,7)	11 (4,17)
	45-64	109 (104,132)	289 (276,352)
	65+	234 (210,256)	524 (472,575)
	Total		824 (763,885)
1969-1988	30-44	4 (1,7)	10 (3,17)
	45-64	103 (89,117)	275 (238,312)
	65+	234 (210,257)	524 (472,576)
	Total		809 (744,873)
1974-1988	30-44	4 (1,8)	10 (3,18)
	45-64	97 (83,111)	257 (220,294)
	65+	238 (214,262)	534 (479,589)
	Total		802 (735,868)

APPENDIX IV(4b continued)

Projected Female Lung Cancer Incidence in 2000  
 Linear Projections of Broad Band Age Specific Rates  
 against Year

Data Years	Age Group	Projected Rate /100,000 (C.I.)	Projected Number of Cases (C.I.)
1964-1988	30-44	3 (1,6)	9 (2,16)
	45-64	159 (138,180)	459 (398,519)
	65+	366 (328,404)	809 (724,894)
	Total		1277 (1172,1381)
1969-1988	30-44	3 (0,6)	8 (1,15)
	45-64	141 (120,163)	407 (346,487)
	65+	366 (324,409)	811 (716,905)
	Total		1226 (1112,1338)
1974-1988	30-44	3 (0,7)	9 (0,17)
	45-64	123 (101,146)	355 (290,420)
	65+	385 (332,437)	851 (734,967)
	Total		1214 (1081,1348)

APPENDIX IV(4b continued)

Projected Female Lung Cancer Incidence in 2010  
 Linear Projections of Broad Band Age Specific Rates  
 against Year

Data Years	Age Group	Projected Rate /100,000 (C.I.)	Projected Number of Cases (C.I.)
1964-1988	30-44	3 (0,5)	7 (0,13)
	45-64	232 (197,268)	763 (647,880)
	65+	573 (501,646)	1305 (1141,1470)
	Total		2075 (1873,2277)
1969-1988	30-44	2 (0,5)	5 (0,12)
	45-64	193 (156,231)	634 (511,758)
	65+	575 (489,661)	1308 (1112,1504)
	Total		1948 (1720,2177)
1974-1988	30-44	3 (0,6)	6 (0,15)
	45-64	155 (118,193)	510 (386,633)
	65+	621 (503,739)	1414 (1145,1683)
	Total		1931 (1634,2227)



APPENDIX VI

Projected Age-Standardised<sup>1</sup> Lung Cancer Incidence Rates for Mersey Region using Models incorporating Smoking Information

Year Lag	Variable	Rate (C.I.)	No. of Cases (C.I.)
<b>Male</b>			
1990 20 yr	Average Tar	128.3 (117.2, 140.4)	1496 (1367, 1638)
1990 15 yr	Average Tar +No. Cigs/Male Smoker	131.0 (121.0, 141.8)	1528 (1411, 1654)
2000 20 yr	Average Tar	90.0 (76.2, 106.4)	1068 (904, 1263)
2000 15 yr	Average Tar +No. Cigs/Male Smoker	106.5 (94.6, 119.8)	1264 (1123, 1422)
2010 20 yr	Average Tar	70.0 (55.6, 88.3)	842 (668, 1061)
2010 15 yr	Average Tar +No. Cigs/Male Smoker	87.0 (73.3, 103.3)	1046 (880, 1242)
<b>Female</b>			
1990 20 yr	No. Cigs/ Female	62.5 (55.3, 70.6)	772 (684, 873)
1990 15 yr	No. Cigs/ Female	60.9 (53.7, 69.1)	753 (664, 854)
2000 20 yr	No. Cigs/ Female	64.6 (57.1, 73.1)	801 (708, 906)
2000 15 yr	No. Cigs/ Female	45.9 (40.7, 51.8)	570 (505, 643)
2010 20 yr	No. Cigs/ Female	45.6 (40.7, 51.1)	564 (504, 632)
2010 15 yr	No. Cigs/ Female	31.3 (27.8, 35.4)	388 (344, 438)

<sup>1</sup> Standardised to the Projected Population for Mersey Region in 2000

APPENDIX VII

Projected Broad-Band Age-Specific Lung Cancer Incidence Rates for Mersey Region using Models incorporating Smoking Information

Year	Age Group	Lag	Rate (C.I.)	No. of Cases (C.I.)
<b>Male</b>				
1990	20 yr	128.3	1496 (117.2, 140.4)	(1367, 1638)
1990	15 yr	Average Tar +No. Cigs/Male Smoker	131.0 (121.0, 141.8)	1528 (1411, 1654)
2000	20 yr	Average Tar	90.0 (76.2, 106.4)	1068 (904, 1263)
2000	15 yr	Average Tar +No. Cigs/Male Smoker	106.5 (94.6, 119.8)	1264 (1123, 1422)
2010	20 yr	Average Tar	70.0 (55.6, 88.3)	842 (668, 1061)
2010	15 yr	Average Tar +No. Cigs/Male Smoker	87.0 (73.3, 103.3)	1046 (880, 1242)
<b>Female</b>				
1990	20 yr	No. Cigs/ Female	62.5 (55.3, 70.6)	772 (684, 873)
1990	15 yr	No. Cigs/ Female	60.9 (53.7, 69.1)	753 (664, 854)
2000	20 yr	No. Cigs/ Female	64.6 (57.1, 73.1)	801 (708, 906)
2000	15 yr	No. Cigs/ Female	45.9 (40.7, 51.8)	570 (505, 643)
2010	20 yr	No. Cigs/ Female	45.6 (40.7, 51.1)	564 (504, 632)
2010	15 yr	No. Cigs/ Female	31.3 (27.8, 35.4)	388 (344, 438)

<sup>1</sup> Standardised to the Projected Population for Mersey Region in 2000

**APPENDIX VIIIa**  
**Projections for Male Lung Cancer Incidence in MRHA**  
**based on Age-Period Cohort Analysis**

Age Group	1991	2001	2011
<b>Age-Specific Incidence Rates (95% Confidence Interval)</b>			
30-34	2.0 (0.3,3.8)	1.4 (0.1,2.7)	0.9 (0,2.0)
35-39	5.2 (1.7,8.7)	3.5 (1.0,5.9)	2.4 (0.4,4.3)
40-44	14.7 (7.0,22.5)	9.6 (4.3,14.9)	6.5 (2.7,10.2)
45-49	39.7 (22.9,56.5)	26.7 (14.7,38.6)	18.0 (9.8,26.1)
50-54	80.2 (49.7,110.8)	59.3 (26.3,92.4)	38.7 (23.8,53.5)
55-59	168.9 (108.2,229.5)	117.6 (59.0,176.2)	79.0 (53.4,104.6)
60-64	346.7 (225.5,467.9)	209.6 (110.1,309.0)	155.0 (55.7,254.2)
65-69	519.8 (339.2,700.4)	348.6 (186.8,510.3)	242.9 (97.7,388.0)
70-74	704.0 (459.7,948.4)	562.6 (303.9,821.3)	340.1 (142.2,538.1)
75-79	770.5 (502.1,1038.9)	694.8 (374.9,1014.6)	466.0 (197.6,734.4)
80-84	907.8 (589.8,1225.7)	811.6 (437.5,1185.8)	648.6 (277.0,1020.2)
<b>Number of Incident Cases (95% Confidence Interval)</b>			
30-34	2 (0,3)	1 (0,2)	1 (0,1)
35-39	4 (1,7)	3 (1,6)	2 (0,3)
40-44	13 (6,19)	8 (4,13)	6 (2,9)
45-49	28 (16,40)	20 (11,29)	17 (9,25)
50-54	52 (32,72)	49 (22,77)	33 (20,45)
55-59	107 (68,145)	79 (39,118)	57 (38,75)
60-64	211 (138,285)	122 (64,180)	117 (42,191)
65-69	278 (181,374)	181 (97,264)	138 (56,220)
70-74	280 (182,377)	247 (133,361)	151 (63,239)
75-79	221 (144,298)	219 (118,320)	158 (67,249)
80-84	147 (96,199)	149 (80,217)	147 (63,231)

**APPENDIX VIIIb**

**Projections for Female Lung Cancer Incidence in MRHA based on Age-Period Cohort Analysis**

Age Group	1991	2001	2011
<b>Age-Specific Incidence Rates (95% Confidence Interval)</b>			
35-39	4.1 (1.7, 6.5)	3.7 (1.5, 5.8)	3.3 (1.2, 5.5)
40-44	9.7 (3.2, 16.3)	8.8 (4.9, 12.7)	8.0 (4.4, 11.5)
45-49	21.7 (13.6, 29.9)	19.8 (12.6, 27.0)	17.9 (11.5, 24.3)
50-54	50.5 (38.0, 62.9)	43.7 (17.1, 70.4)	39.6 (27.1, 52.0)
55-59	87.1 (71.4, 102.8)	83.0 (56.4, 109.7)	75.7 (53.3, 98.2)
60-64	171.2 (149.5, 192.9)	156.4 (123.1, 189.7)	135.5 (54.6, 216.4)
65-69	254.4 (228.0, 280.9)	217.1 (182.3, 251.9)	206.9 (142.6, 271.3)
70-74	288.0 (258.7, 317.3)	347.1 (305.1, 389.1)	317.1 (250.8, 383.4)
75-79	271.8 (242.1, 301.4)	432.5 (385.0, 480.0)	369.0 (308.7, 429.2)
80-84	243.0 (210.9, 275.0)	422.0 (370.1, 474.0)	508.7 (437.9, 579.5)
<b>Number of Incident Cases (95% Confidence Interval)</b>			
35-39	3 (1, 5)	4 (1, 6)	3 (1, 4)
40-44	8 (3, 14)	8 (4, 11)	7 (4, 10)
45-49	16 (10, 22)	15 (10, 21)	17 (11, 23)
50-54	34 (26, 42)	36 (14, 58)	34 (23, 44)
55-59	56 (46, 66)	57 (38, 75)	56 (39, 72)
60-64	113 (99, 127)	97 (76, 118)	105 (42, 167)
65-69	162 (145, 179)	123 (103, 143)	126 (87, 166)
70-74	158 (141, 174)	187 (164, 209)	166 (131, 200)
75-79	129 (115, 143)	199 (177, 221)	159 (133, 185)
80-84	84 (73, 95)	140 (123, 158)	175 (151, 200)



**APPENDIX IX(1a)**

**Scenario 9.1: Projections for Male Lung Cancer Incidence in MRHA based on Extrapolation of Average Tar Content and Percentage of Males who Smoke**

Age Group	1991	2001	2011
<b>Age-Specific Incidence Rates (95% Confidence Interval)</b>			
30-34	2.5 (0.9,4.1)	2.0 (0.6,3.4)	1.6 (0.2,2.9)
35-39	6.6 (3.8,9.3)	5.2 (3.0,7.5)	4.1 (2.0,6.3)
40-44	17.9 (13.5,22.3)	13.9 (10.0,17.9)	11.1 (7.5,14.6)
45-49	49.4 (41.2,57.6)	37.3 (30.0,44.7)	29.8 (23.3,36.3)
50-54	99.2 (86.8,111.6)	78.2 (66.9,89.4)	60.8 (50.0,71.7)
55-59	191.2 (173.3,209.2)	154.8 (136.9,172.8)	116.7 (99.1,134.2)
60-64	334.1 (309.2,358.9)	264.5 (238.7,290.3)	208.0 (182.4,233.5)
65-69	490.0 (456.5,523.6)	393.7 (360.6,426.9)	318.0 (283.7,352.4)
70-74	679.8 (632.7,726.8)	540.6 (499.3,581.8)	427.0 (384.5,469.5)
75-79	740.2 (684.5,796.0)	656.1 (604.2,707.9)	525.9 (476.0,575.8)
80-84	869.4 (789.1,949.6)	785.6 (710.1,861.2)	623.3 (560.1,686.4)

**Number of Incident Cases (95% Confidence Interval)**

30-34	2 (1,4)	2 (1,3)	1 (0,2)
35-39	5 (3,7)	5 (3,7)	3 (2,5)
40-44	15 (12,19)	12 (9,16)	10 (7,13)
45-49	35 (29,41)	28 (23,34)	29 (23,35)
50-54	65 (57,73)	65 (55,74)	51 (42,60)
55-59	121 (110,132)	106 (91,115)	84 (71,97)
60-64	204 (189,219)	154 (139,169)	157 (137,176)
65-69	262 (244,280)	204 (187,221)	181 (161,200)
70-74	270 (251,289)	237 (219,255)	190 (171,209)
75-79	212 (196,228)	207 (190,223)	178 (161,195)
80-84	141 (128,154)	144 (130,158)	141 (127,155)

**APPENDIX IX(1b)**

**Scenario 9.1: Projections for Female Lung Cancer Incidence in MRHA based on Extrapolation of Average Tar Content and Percentage of Females who Smoke**

Age Group	1991	2001	2011
<b>Age-Specific Incidence Rates (95% Confidence Interval)</b>			
35-39	3.5 (1.5, 5.5)	3.1 (1.3, 4.9)	2.7 (0.9, 4.6)
40-44	9.5 (6.2, 12.8)	7.7 (4.6, 10.7)	6.9 (3.9, 10.0)
45-49	21.7 (16.2, 27.3)	17.9 (12.5, 23.2)	15.8 (10.6, 21.1)
50-54	46.7 (37.9, 55.4)	42.0 (33.0, 50.9)	33.8 (24.5, 43.1)
55-59	91.6 (78.7, 104.5)	82.6 (68.1, 97.1)	67.8 (52.1, 83.5)
60-64	161.4 (143.9, 179.0)	144.5 (123.6, 165.3)	129.6 (105.3, 153.9)
65-69	250.1 (225.1, 275.1)	225.5 (198.0, 253.1)	203.0 (170.1, 235.8)
70-74	280.3 (251.1, 309.5)	324.8 (289.8, 359.8)	290.3 (248.9, 331.7)
75-79	263.5 (234.0, 293.0)	419.9 (374.9, 465.0)	378.2 (328.8, 427.6)
80-84	234.1 (202.4, 265.8)	403.4 (351.7, 455.1)	467.0 (405.3, 528.6)

**Number of Incident Cases (95% Confidence Interval)**

35-39	3 (1, 4)	3 (1, 5)	2 (1, 3)
40-44	8 (5, 11)	7 (4, 9)	6 (3, 9)
45-49	16 (12, 20)	14 (10, 18)	15 (10, 20)
50-54	31 (26, 37)	35 (27, 42)	29 (21, 37)
55-59	59 (51, 67)	56 (46, 66)	50 (39, 62)
60-64	107 (95, 118)	90 (77, 103)	100 (81, 119)
65-69	159 (143, 175)	128 (112, 144)	124 (104, 144)
70-74	153 (137, 169)	175 (156, 193)	152 (130, 173)
75-79	125 (111, 139)	193 (173, 214)	163 (142, 185)
80-84	81 (70, 92)	134 (117, 151)	161 (140, 182)

APPENDIX IX(2a)

Scenario 9.2: Projections for Male Lung Cancer Incidence in MRHA based on Extrapolation of Average Tar Content and the Health of the Nation Target for the Percentage of Males who Smoke

Age Group	1991	2001	2011
<b>Age-Specific Incidence Rates (95% Confidence Interval)</b>			
30-34	2.5 (0.9,4.1)	1.7 (0.4,2.9)	1.4 (0.1,2.7)
35-39	6.6 (3.8,9.3)	4.3 (2.3,6.3)	3.6 (1.6,5.6)
40-44	17.9 (13.5,22.3)	13.9 (10.0,17.9)	9.3 (6.1,12.6)
45-49	49.4 (41.2,57.6)	37.3 (30.0,44.7)	24.3 (18.4,30.2)
50-54	99.2 (86.8,111.6)	78.2 (66.9,89.4)	60.8 (50.0,71.7)
55-59	191.2 (173.3,209.2)	154.8 (136.9,172.8)	116.7 (99.1,134.2)
60-64	334.1 (309.2,358.9)	264.5 (238.7,290.3)	208.0 (182.4,233.5)
65-69	490.0 (456.5,523.6)	393.7 (360.6,426.9)	318.0 (283.7,352.4)
70-74	679.8 (632.7,726.8)	540.6 (499.3,581.8)	427.0 (384.5,469.5)
75-79	740.2 (684.5,796.0)	656.1 (604.2,707.9)	525.9 (476.0,575.8)
80-84	869.4 (789.1,949.6)	785.6 (710.1,861.2)	623.3 (560.1,686.4)

**Number of Incident Cases (95% Confidence Interval)**

30-34	2 (1,4)	2 (0,3)	1 (0,2)
35-39	5 (3,7)	4 (2,6)	3 (1,4)
40-44	15 (12,19)	12 (9,16)	9 (6,12)
45-49	35 (29,41)	28 (23,34)	24 (18,29)
50-54	65 (57,73)	65 (55,74)	51 (42,60)
55-59	121 (110,132)	106 (91,115)	84 (71,97)
60-64	204 (189,219)	154 (139,169)	157 (137,176)
65-69	262 (244,280)	204 (187,221)	181 (161,200)
70-74	270 (251,289)	237 (219,255)	190 (171,209)
75-79	212 (196,228)	207 (190,223)	178 (161,195)
80-84	141 (128,154)	144 (130,158)	141 (127,155)

APPENDIX IX(2b)

**Scenario 9.2: Projections for Female Lung Cancer Incidence in MRHA based on Extrapolation of Average Tar Content and the Health of the Nation Target for the Percentage of Females who Smoke**

Age Group	1991	2001	2011
<b>Age-Specific Incidence Rates (95% Confidence Interval)</b>			
35-39	3.5 (1.5, 5.5)	2.1 (0.7, 3.5)	1.9 (0.4, 3.4)
40-44	9.5 (6.2, 12.8)	7.7 (4.6, 10.7)	4.7 (2.3, 7.2)
45-49	21.7 (16.2, 27.3)	17.9 (12.5, 23.2)	10.6 (6.4, 14.7)
50-54	46.7 (37.9, 55.4)	42.0 (33.0, 50.9)	33.8 (24.5, 43.1)
55-59	91.6 (78.7, 104.5)	82.6 (68.1, 97.1)	67.8 (52.1, 83.5)
60-64	161.4 (143.9, 179.0)	144.5 (123.6, 165.3)	129.6 (105.3, 153.9)
65-69	250.1 (225.1, 275.1)	225.5 (198.0, 253.1)	203.0 (170.1, 235.8)
70-74	280.3 (251.1, 309.5)	324.8 (289.8, 359.8)	290.3 (248.9, 331.7)
75-79	263.5 (234.0, 293.0)	419.9 (374.9, 465.0)	378.2 (328.8, 427.6)
80-84	234.1 (202.4, 265.8)	403.4 (351.7, 455.1)	467.0 (405.3, 528.6)
<b>Number of Incident Cases (95% Confidence Interval)</b>			
35-39	3 (1, 4)	2 (1, 3)	1 (0, 3)
40-44	8 (5, 11)	7 (4, 9)	4 (2, 6)
45-49	16 (12, 20)	14 (10, 18)	10 (6, 14)
50-54	31 (26, 37)	35 (27, 42)	29 (21, 37)
55-59	59 (51, 67)	56 (46, 66)	50 (39, 62)
60-64	107 (95, 118)	90 (77, 103)	100 (81, 119)
65-69	159 (143, 175)	128 (112, 144)	124 (104, 144)
70-74	153 (137, 169)	175 (156, 193)	152 (130, 173)
75-79	125 (111, 139)	193 (173, 214)	163 (142, 185)
80-84	81 (70, 92)	134 (117, 151)	161 (140, 182)



APPENDIX IX(3a)

**Scenario 9.3: Projections for Male Lung Cancer Incidence in MRHA based on 1986 Average Tar Content and Extrapolation of the Percentage of Males who Smoke**

Age Group	1991	2001	2011
<b>Age-Specific Incidence Rates (95% Confidence Interval)</b>			
30-34	2.5 (0.9,4.1)	2.1 (0.7,3.6)	2.1 (0.5,3.6)
35-39	6.6 (3.8,9.3)	5.8 (3.4,8.1)	5.5 (3.0,7.9)
40-44	17.9 (13.5,22.3)	15.4 (11.2,19.5)	14.7 (10.7,18.6)
45-49	49.4 (41.2,57.6)	41.1 (33.6,48.7)	39.4 (32.6,46.2)
50-54	99.2 (86.8,111.6)	86.0 (74.9,97.2)	80.5 (69.6,91.4)
55-59	191.2 (173.3,209.2)	170.5 (152.8,188.1)	154.3 (137.5,171.2)
60-64	334.1 (309.2,358.9)	291.2 (266.1,316.2)	275.0 (252.0,298.1)
65-69	490.0 (456.5,523.6)	433.5 (400.9,466.0)	420.6 (389.4,451.9)
70-74	679.8 (632.7,726.8)	595.1 (553.4,636.8)	564.7 (524.6,604.9)
75-79	740.2 (684.5,796.0)	722.2 (667.5,776.9)	695.5 (643.8,747.3)
80-84	869.4 (789.1,949.6)	864.9 (786.6,943.1)	824.3 (751.4,897.2)

**Number of Incident Cases (95% Confidence Interval)**

30-34	2 (1,4)	2 (1,3)	1 (0,3)
35-39	5 (3,7)	6 (3,8)	4 (2,6)
40-44	15 (12,19)	13 (10,17)	13 (10,17)
45-49	35 (29,41)	31 (26,37)	38 (32,45)
50-54	65 (57,73)	71 (62,81)	68 (58,77)
55-59	121 (110,132)	114 (102,126)	111 (99,123)
60-64	204 (189,219)	170 (155,185)	207 (190,224)
65-69	262 (244,280)	225 (208,242)	239 (221,257)
70-74	270 (251,289)	261 (243,280)	251 (234,269)
75-79	212 (196,228)	228 (210,245)	235 (218,253)
80-84	141 (128,154)	159 (144,173)	186 (170,203)

**APPENDIX IX(3b)**

**Scenario 9.3: Projections for Female Lung Cancer Incidence in MRHA based on 1986 Average Tar Content and and Extrapolation of the Percentage of Females who Smoke**

Age Group	1991	2001	2011
<b>Age-Specific Incidence Rates (95% Confidence Interval)</b>			
35-39	3.5 (1.5, 5.5)	3.3 (1.5, 5.1)	3.1 (1.2, 5.1)
40-44	9.5 (6.2, 12.8)	8.0 (5.0, 11.1)	8.0 (5.0, 11.0)
45-49	21.7 (16.2, 27.3)	18.8 (13.5, 24.0)	18.3 (13.5, 23.1)
50-54	46.7 (37.9, 55.4)	44.1 (35.6, 52.6)	39.0 (31.0, 47.0)
55-59	91.6 (78.7, 104.5)	86.7 (73.3, 100.1)	78.2 (65.6, 90.8)
60-64	161.4 (143.9, 179.0)	151.7 (132.8, 170.6)	149.5 (131.7, 167.4)
65-69	250.1 (225.1, 275.1)	236.9 (211.2, 262.6)	234.1 (209.1, 259.2)
70-74	280.3 (251.1, 309.5)	341.1 (306.8, 375.4)	334.8 (300.9, 368.8)
75-79	263.5 (234.0, 293.0)	441.0 (393.4, 488.7)	436.2 (388.7, 483.7)
80-84	234.1 (202.4, 265.8)	423.7 (371.5, 475.9)	538.6 (470.0, 607.1)

**Number of Incident Cases (95% Confidence Interval)**

35-39	3 (1, 4)	3 (1, 5)	2 (1, 4)
40-44	8 (5, 11)	7 (4, 10)	7 (4, 10)
45-49	16 (12, 20)	15 (10, 19)	18 (13, 22)
50-54	31 (26, 37)	37 (30, 44)	33 (26, 40)
55-59	59 (51, 67)	59 (50, 68)	58 (48, 67)
60-64	107 (95, 118)	94 (83, 106)	115 (102, 129)
65-69	159 (143, 175)	134 (120, 149)	143 (128, 158)
70-74	153 (137, 169)	183 (165, 202)	175 (157, 193)
75-79	125 (111, 139)	203 (181, 225)	188 (168, 209)
80-84	81 (70, 92)	141 (124, 158)	186 (162, 209)

**APPENDIX IX(4a)**

**Scenario 9.4: Projections for Male Lung Cancer Incidence in MRHA based on 1986 Average Tar Content and the Health Target for the Percentage of Males who Smoke**

Age Group	1991	2001	2011
<b>Age-Specific Incidence Rates (95% Confidence Interval)</b>			
30-34	2.5 (0.9,4.1)	1.8 (0.5,3.1)	1.8 (0.4,3.3)
35-39	6.6 (3.8,9.3)	4.7 (2.6,6.8)	4.7 (2.4,7.0)
40-44	17.9 (13.5,22.3)	15.4 (11.2,19.5)	12.3 (8.7,15.9)
45-49	49.4 (41.2,57.6)	41.1 (33.6,48.7)	32.2 (25.8,38.5)
50-54	99.2 (86.8,111.6)	86.0 (74.9,97.2)	80.5 (69.6,91.4)
55-59	191.2 (173.3,209.2)	170.5 (152.8,188.1)	154.3 (137.5,171.2)
60-64	334.1 (309.2,358.9)	291.2 (266.1,316.2)	275.0 (252.0,298.1)
65-69	490.0 (456.5,523.6)	433.5 (400.9,466.0)	420.6 (389.4,451.9)
70-74	679.8 (632.7,726.8)	595.1 (553.4,636.8)	564.7 (524.6,604.9)
75-79	740.2 (684.5,796.0)	722.2 (667.5,776.9)	695.5 (643.8,747.3)
80-84	869.4 (789.1,949.6)	864.9 (786.6,943.1)	824.3 (751.4,897.2)

**Number of Incident Cases (95% Confidence Interval)**

30-34	2 (1,4)	2 (0,3)	1 (0,2)
35-39	5 (3,7)	5 (3,7)	4 (2,6)
40-44	15 (12,19)	13 (10,17)	11 (8,15)
45-49	35 (29,41)	31 (26,37)	31 (25,37)
50-54	65 (57,73)	71 (62,81)	68 (58,77)
55-59	121 (110,132)	114 (102,126)	111 (99,123)
60-64	204 (189,219)	170 (155,185)	207 (190,224)
65-69	262 (244,280)	225 (208,242)	239 (221,257)
70-74	270 (251,289)	261 (243,280)	251 (234,269)
75-79	212 (196,228)	228 (210,245)	235 (218,253)
80-84	141 (128,154)	159 (144,173)	186 (170,203)

APPENDIX IX(4b)

**Scenario 9.4: Projections for Female Lung Cancer Incidence in MRHA based on 1986 Average Tar Content and the Health of the Nation Targets for the Percentage of Females who Smoke**

Age Group	1991	2001	2011
<b>Age-Specific Incidence Rates (95% Confidence Interval)</b>			
35-39	3.5 (1.5,5.5)	2.1 (0.7,3.6)	2.2 (0.6,3.8)
40-44	9.5 (6.2,12.8)	8.0 (5.0,11.1)	5.5 (3.0,7.9)
45-49	21.7 (16.2,27.3)	18.8 (13.5,24.0)	12.2 (8.2,16.1)
50-54	46.7 (37.9,55.4)	44.1 (35.6,52.6)	39.0 (31.0,47.0)
55-59	91.6 (78.7,104.5)	86.7 (73.3,100.1)	78.2 (65.6,90.8)
60-64	161.4 (143.9,179.0)	151.7 (132.8,170.6)	149.5 (131.7,167.4)
65-69	250.1 (225.1,275.1)	236.9 (211.2,262.6)	234.1 (209.1,259.2)
70-74	280.3 (251.1,309.5)	341.1 (306.8,375.4)	334.8 (300.9,368.8)
75-79	263.5 (234.0,293.0)	441.0 (393.4,488.7)	436.2 (388.7,483.7)
80-84	234.1 (202.4,265.8)	423.7 (371.5,475.9)	538.6 (470.0,607.1)
<b>Number of Incident Cases (95% Confidence Interval)</b>			
35-39	3 (1,4)	2 (1,4)	2 (0,3)
40-44	8 (5,11)	7 (4,10)	5 (3,7)
45-49	16 (12,20)	15 (10,19)	12 (8,16)
50-54	31 (26,37)	37 (30,44)	33 (26,40)
55-59	59 (51,67)	59 (50,68)	58 (48,67)
60-64	107 (95,118)	94 (83,106)	115 (102,129)
65-69	159 (143,175)	134 (120,149)	143 (128,158)
70-74	153 (137,169)	183 (165,202)	175 (157,193)
75-79	125 (111,139)	203 (181,225)	188 (168,209)
80-84	81 (70,92)	141 (124,158)	186 (162,209)



APPENDIX IX(5a)

Scenario 9.5: Projections for Male Lung Cancer Incidence in MRHA based on Extrapolation of Average Tar Content and the Number of Cigarettes Smoked per Male

Age Group	1991	2001	2011
<b>Age-Specific Incidence Rates (95% Confidence Interval)</b>			
30-34	2.5 (0.9,4.0)	1.8 (0.5,3.2)	1.4 (0.1,2.6)
35-39	6.7 (3.9,9.5)	5.0 (2.9,7.2)	3.8 (1.7,5.8)
40-44	18.9 (14.3,23.4)	14.1 (10.1,18.1)	10.5 (7.1,14.0)
45-49	52.3 (43.8,60.7)	39.0 (31.5,46.6)	29.3 (22.8,35.8)
50-54	102.1 (89.5,114.6)	83.8 (72.2,95.3)	62.4 (51.3,73.4)
55-59	196.6 (178.4,214.7)	165.0 (146.5,183.4)	122.8 (104.8,140.9)
60-64	337.5 (312.5,362.5)	273.2 (247.0,299.3)	223.7 (197.3,250.1)
65-69	481.2 (448.1,514.3)	403.6 (370.1,437.1)	337.9 (302.6,373.2)
70-74	677.2 (630.2,724.2)	543.8 (502.4,585.2)	439.0 (396.0,482.1)
75-79	736.7 (681.1,792.3)	642.8 (591.7,693.9)	537.8 (487.3,588.3)
80-84	861.9 (782.0,941.7)	778.3 (703.2,853.4)	623.3 (560.2,686.5)
<b>Number of Incident Cases (95% Confidence Interval)</b>			
30-34	2 (1,3)	2 (0,3)	1 (0,2)
35-39	5 (3,7)	5 (3,7)	3 (1,5)
40-44	16 (12,20)	12 (9,16)	10 (6,13)
45-49	37 (31,43)	30 (24,35)	28 (22,35)
50-54	67 (59,75)	69 (60,79)	52 (43,62)
55-59	124 (113,136)	110 (98,123)	88 (75,101)
60-64	206 (191,221)	159 (144,175)	168 (149,188)
65-69	257 (239,275)	209 (192,227)	192 (172,212)
70-74	269 (250,288)	239 (221,257)	195 (176,215)
75-79	211 (195,227)	202 (186,219)	182 (165,199)
80-84	140 (127,153)	143 (129,157)	141 (127,155)

**APPENDIX IX(5b)**

**Scenario 9.5: Projections for Female Lung Cancer Incidence in MRHA based on Extrapolation of Average Tar Content and the Number of Cigarettes Smoked per Female**

Age Group	1991	2001	2011
<b>Age-Specific Incidence Rates (95% Confidence Interval)</b>			
35-39	3.8 (1.7,5.9)	2.8 (1.1,4.5)	2.1 (0.5,3.7)
40-44	11.5 (7.9,15.2)	7.8 (4.7,10.9)	6.0 (3.2,8.9)
45-49	26.1 (20.0,32.2)	20.9 (15.1,26.7)	15.4 (10.3,20.6)
50-54	51.5 (42.3,60.6)	53.9 (43.7,64.2)	36.4 (26.7,46.2)
55-59	98.3 (85.0,111.6)	102.5 (86.4,118.7)	82.0 (64.5,99.5)
60-64	172.6 (154.3,190.9)	163.9 (141.8,186.1)	171.4 (143.1,199.8)
65-69	232.5 (208.9,256.1)	241.3 (212.7,269.9)	251.1 (214.1,288.2)
70-74	274.7 (245.8,303.5)	339.7 (303.6,375.9)	322.0 (278.1,365.8)
75-79	256.9 (227.9,285.9)	386.5 (344.4,428.5)	400.3 (349.0,451.5)
80-84	224.9 (194.1,255.7)	386.2 (336.2,436.2)	476.6 (413.7,539.5)

**Number of Incident Cases (95% Confidence Interval)**

35-39	3 (1,5)	3 (1,4)	2 (0,3)
40-44	10 (7,13)	7 (4,10)	5 (3,8)
45-49	19 (14,23)	16 (12,21)	15 (10,20)
50-54	35 (28,41)	45 (36,53)	31 (23,39)
55-59	64 (55,72)	70 (59,81)	60 (48,73)
60-64	114 (102,126)	102 (88,116)	132 (110,154)
65-69	148 (133,163)	137 (121,153)	153 (131,176)
70-74	150 (134,166)	183 (163,202)	168 (145,191)
75-79	122 (108,136)	178 (159,197)	173 (151,195)
80-84	78 (67,88)	129 (112,145)	164 (143,186)

**APPENDIX IX(6a)**

**Scenario 9.6: Projections for Male Lung Cancer Incidence in MRHA based on Extrapolation of Average Tar Content and the Health of the Nation Target for the Number of Cigarettes Smoked per Male**

Age Group	1991	2001	2011
<b>Age-Specific Incidence Rates (95% Confidence Interval)</b>			
30-34	2.5 (0.9,4.0)	1.7 (0.4,3.0)	1.4 (0.1,2.7)
35-39	6.7 (3.9,9.5)	4.5 (2.5,6.6)	3.7 (1.7,5.8)
40-44	18.9 (14.3,23.4)	14.1 (10.1,18.1)	10.0 (6.5,13.3)
45-49	52.3 (43.8,60.7)	39.0 (31.5,46.6)	26.3 (20.1,32.4)
50-54	102.1 (89.5,114.6)	83.8 (72.2,95.3)	62.4 (51.3,73.4)
55-59	196.6 (178.4,214.7)	165.0 (146.5,183.4)	122.8 (104.8,140.9)
60-64	337.5 (312.5,362.5)	273.2 (247.0,299.3)	223.7 (197.3,250.1)
65-69	481.2 (448.1,514.3)	403.6 (370.1,437.1)	337.9 (302.6,373.2)
70-74	677.2 (630.2,724.2)	543.8 (502.4,585.2)	439.0 (396.0,482.1)
75-79	736.7 (681.1,792.3)	642.8 (591.7,693.9)	537.8 (487.3,588.3)
80-84	861.9 (782.0,941.7)	778.3 (703.2,853.4)	623.3 (560.2,686.5)

**Number of Incident Cases (95% Confidence Interval)**

30-34	2 (1,3)	2 (0,3)	1 (0,2)
35-39	5 (3,7)	4 (2,6)	3 (1,5)
40-44	16 (12,20)	12 (9,16)	9 (6,12)
45-49	37 (31,43)	30 (24,35)	25 (19,31)
50-54	67 (59,75)	69 (60,79)	52 (43,62)
55-59	124 (113,136)	110 (98,123)	88 (75,101)
60-64	206 (191,221)	159 (144,175)	168 (149,188)
65-69	257 (239,275)	209 (192,227)	192 (172,212)
70-74	269 (250,288)	239 (221,257)	195 (176,215)
75-79	211 (195,227)	202 (186,219)	182 (165,199)
80-84	140 (127,153)	143 (129,157)	141 (127,155)

APPENDIX IX(6b)

**Scenario 9.6: Projections for Female Lung Cancer Incidence in MRHA based on Extrapolation of Average Tar Content and the Health of the Nation Target for the Number of Cigarettes Smoked per Female**

Age Group	1991	2001	2011
<b>Age-Specific Incidence Rates (95% Confidence Interval)</b>			
35-39	3.8 (1.7, 5.9)	2.0 (0.6, 3.4)	1.7 (0.3, 3.1)
40-44	11.5 (7.9, 15.2)	7.8 (4.7, 10.9)	4.6 (2.2, 7.1)
45-49	26.1 (20.0, 32.2)	20.9 (15.1, 26.7)	11.1 (6.8, 15.5)
50-54	51.5 (42.3, 60.6)	53.9 (43.7, 64.2)	36.4 (26.7, 46.2)
55-59	98.3 (85.0, 111.6)	102.5 (86.4, 118.7)	82.0 (64.5, 99.5)
60-64	172.6 (154.3, 190.9)	163.9 (141.8, 186.1)	171.4 (143.1, 199.8)
65-69	232.5 (208.9, 256.1)	241.3 (212.7, 269.9)	251.1 (214.1, 288.2)
70-74	274.7 (245.8, 303.5)	339.7 (303.6, 375.9)	322.0 (278.1, 365.8)
75-79	256.9 (227.9, 285.9)	386.5 (344.4, 428.5)	400.3 (349.0, 451.5)
80-84	224.9 (194.1, 255.7)	386.2 (336.2, 436.2)	476.6 (413.7, 539.5)
<b>Number of Incident Cases (95% Confidence Interval)</b>			
35-39	3 (1, 5)	2 (1, 3)	1 (0, 2)
40-44	10 (7, 13)	7 (4, 10)	4 (2, 6)
45-49	19 (14, 23)	16 (12, 21)	11 (7, 15)
50-54	35 (28, 41)	45 (36, 53)	31 (23, 39)
55-59	64 (55, 72)	70 (59, 81)	60 (48, 73)
60-64	114 (102, 126)	102 (88, 116)	132 (110, 154)
65-69	148 (133, 163)	137 (121, 153)	153 (131, 176)
70-74	150 (134, 166)	183 (163, 202)	168 (145, 191)
75-79	122 (108, 136)	178 (159, 197)	173 (151, 195)
80-84	78 (67, 88)	129 (112, 145)	164 (143, 186)



**APPENDIX IX(7a)**

**Scenario 9.7: Projections for Male Lung Cancer Incidence in MRHA based on the 1986 Average Tar Content and Extrapolation of the Number of Cigarettes Smoked per Male**

Age Group	1991	2001	2011
<b>Age-Specific Incidence Rates (95% Confidence Interval)</b>			
30-34	2.5 (0.9,4.0)	2.0 (0.6,3.4)	1.8 (0.4,3.3)
35-39	6.7 (3.9,9.5)	5.6 (3.3,7.9)	5.1 (2.7,7.4)
40-44	18.9 (14.4,23.4)	15.6 (11.4,19.8)	14.1 (10.2,18.0)
45-49	52.3 (43.8,60.7)	43.2 (35.4,50.9)	39.3 (32.5,46.2)
50-54	102.1 (89.5,114.6)	92.7 (81.2,104.2)	83.7 (72.6,94.8)
55-59	196.6 (178.4,214.7)	182.5 (164.4,200.6)	164.8 (147.6,182.0)
60-64	337.5 (312.5,362.5)	302.2 (276.8,327.6)	300.1 (276.6,323.6)
65-69	481.2 (448.1,514.3)	446.5 (413.5,479.5)	453.3 (420.9,485.8)
70-74	677.2 (630.2,724.2)	601.5 (559.5,643.6)	589.0 (547.8,630.2)
75-79	736.7 (681.1,792.3)	711.1 (657.1,765.1)	721.5 (667.9,775.1)
80-84	861.9 (782.0,941.7)	861.0 (782.9,939.0)	836.3 (761.7,910.8)

**Number of Incident Cases (95% Confidence Interval)**

30-34	2 (1,3)	2 (1,3)	1 (0,2)
35-39	5 (3,7)	5 (3,8)	4 (2,6)
40-44	16 (12,20)	13 (10,17)	13 (9,17)
45-49	37 (31,43)	33 (27,39)	38 (31,45)
50-54	67 (59,75)	77 (67,86)	70 (61,80)
55-59	124 (113,136)	122 (110,134)	118 (106,131)
60-64	206 (191,221)	176 (162,191)	226 (208,244)
65-69	257 (239,275)	232 (214,249)	258 (239,276)
70-74	269 (250,288)	264 (246,283)	262 (244,280)
75-79	211 (195,227)	224 (207,241)	244 (226,262)
80-84	140 (127,153)	158 (144,172)	189 (172,206)

APPENDIX IX(7b)

**Scenario 9.7: Projections for Female Lung Cancer Incidence in MRHA based on 1986 Average Tar Content and and Extrapolation of the Number of Cigarettes Smoked per Female**

Age Group	1991	2001	2011
<b>Age-Specific Incidence Rates (95% Confidence Interval)</b>			
35-39	3.8 (1.7, 5.9)	3.0 (1.3, 4.8)	2.7 (0.9, 4.5)
40-44	11.5 (7.9, 15.2)	8.5 (5.4, 11.7)	8.0 (4.8, 10.8)
45-49	26.1 (20.0, 32.2)	22.8 (17.0, 28.6)	19.9 (14.9, 25.0)
50-54	51.5 (42.3, 60.6)	58.9 (49.1, 68.7)	47.0 (38.3, 55.7)
55-59	98.3 (85.0, 111.6)	111.9 (96.7, 127.1)	105.8 (91.4, 120.2)
60-64	172.6 (154.3, 190.9)	178.9 (158.3, 199.6)	221.3 (197.7, 244.8)
65-69	232.5 (208.9, 256.1)	263.4 (235.5, 291.3)	324.1 (289.3, 358.9)
70-74	274.7 (245.8, 303.5)	370.8 (333.2, 408.5)	415.5 (371.7, 459.3)
75-79	256.9 (227.9, 285.9)	421.9 (376.2, 467.5)	516.5 (456.1, 577.0)
80-84	224.9 (194.1, 255.7)	421.6 (369.4, 473.8)	615.1 (529.9, 700.2)

**Number of Incident Cases (95% Confidence Interval)**

35-39	3 (1, 5)	3 (1, 5)	2 (1, 3)
40-44	10 (7, 13)	7 (5, 10)	7 (4, 10)
45-49	19 (14, 23)	18 (13, 22)	19 (14, 24)
50-54	35 (28, 41)	49 (40, 57)	40 (33, 47)
55-59	64 (55, 72)	76 (66, 87)	78 (67, 89)
60-64	114 (102, 126)	111 (98, 124)	171 (153, 189)
65-69	148 (133, 164)	149 (133, 165)	198 (177, 219)
70-74	150 (135, 166)	199 (179, 219)	217 (194, 240)
75-79	122 (108, 136)	194 (173, 215)	223 (197, 249)
80-84	78 (67, 88)	140 (123, 158)	212 (183, 241)

**APPENDIX IX(8a)**

**Scenario 9.8: Projections for Male Lung Cancer Incidence in MRHA based on the 1986 Average Tar Content and the Health of the Nation Target for the Number of Cigarettes Smoked per Male**

Age Group	1991	2001	2011
<b>Age-Specific Incidence Rates (95% Confidence Interval)</b>			
30-34	2.5 (0.9,4.0)	1.9 (0.6,3.3)	1.9 (0.4,3.4)
35-39	6.7 (3.9,9.5)	5.0 (2.8,7.2)	5.0 (2.6,7.4)
40-44	18.9 (14.4,23.4)	15.6 (11.4,19.8)	13.3 (9.5,17.1)
45-49	52.3 (43.8,60.7)	43.2 (35.4,50.9)	35.2 (28.6,41.8)
50-54	102.1 (89.5,114.6)	92.7 (81.2,104.2)	83.7 (72.6,94.8)
55-59	196.6 (178.4,214.7)	182.5 (164.4,200.6)	164.8 (147.6,182.0)
60-64	337.5 (312.5,362.5)	302.2 (276.8,327.6)	300.1 (276.6,323.6)
65-69	481.2 (448.1,514.3)	446.5 (413.5,479.5)	453.3 (420.9,485.8)
70-74	677.2 (630.2,724.2)	601.5 (559.5,643.6)	589.0 (547.8,630.2)
75-79	736.7 (681.1,792.3)	711.1 (657.1,765.1)	721.5 (667.9,775.1)
80-84	861.9 (782.0,941.7)	861.0 (782.9,939.0)	836.3 (761.7,910.8)

**Number of Incident Cases (95% Confidence Interval)**

30-34	2 (1,3)	2 (1,3)	1 (0,2)
35-39	5 (3,7)	5 (3,7)	4 (2,6)
40-44	16 (12,20)	13 (10,17)	12 (9,16)
45-49	37 (31,43)	33 (27,39)	34 (28,40)
50-54	67 (59,75)	77 (67,86)	70 (61,80)
55-59	124 (113,136)	122 (110,134)	118 (106,131)
60-64	206 (191,221)	176 (162,191)	226 (208,244)
65-69	257 (239,275)	232 (214,249)	258 (239,276)
70-74	269 (250,288)	264 (246,283)	262 (244,280)
75-79	211 (195,227)	224 (207,241)	244 (226,262)
80-84	140 (127,153)	158 (144,172)	189 (172,206)

APPENDIX IX(8b)

Scenario 9.8: Projections for Female Lung Cancer Incidence in MRHA based on 1986 Average Tar Content and the Health of the Nation Target for the Number of Cigarettes Smoked per Female

Age Group	1991	2001	2011
Age-Specific Incidence Rates (95% Confidence Interval)			
35-39	3.8 (1.7, 5.9)	2.2 (0.7, 3.6)	2.2 (0.6, 3.8)
40-44	11.5 (7.9, 15.2)	8.5 (5.4, 11.7)	6.0 (3.4, 8.6)
45-49	26.1 (20.0, 32.2)	22.8 (17.0, 28.6)	14.4 (10.1, 18.7)
50-54	51.5 (42.3, 60.6)	59.9 (49.1, 68.7)	47.0 (38.3, 55.7)
55-59	98.3 (85.0, 111.6)	111.9 (96.7, 127.1)	105.8 (91.4, 120.2)
60-64	172.6 (154.3, 190.9)	178.9 (158.3, 199.6)	221.3 (197.7, 244.8)
65-69	232.5 (208.9, 256.1)	263.4 (235.5, 291.3)	324.1 (289.3, 358.9)
70-74	274.7 (245.8, 303.5)	370.8 (333.2, 408.5)	415.5 (371.7, 459.3)
75-79	256.9 (227.9, 285.9)	421.9 (376.2, 467.5)	516.5 (456.1, 577.0)
80-84	224.9 (194.1, 255.7)	421.6 (369.4, 473.8)	615.1 (529.9, 700.2)

Number of Incident Cases (95% Confidence Interval)

35-39	3 (1, 5)	2 (1, 4)	2 (0, 3)
40-44	10 (7, 13)	7 (5, 10)	5 (3, 8)
45-49	19 (14, 23)	18 (13, 22)	14 (10, 18)
50-54	35 (28, 41)	49 (40, 57)	40 (33, 47)
55-59	64 (55, 72)	76 (66, 87)	78 (67, 89)
60-64	114 (102, 126)	111 (98, 124)	171 (153, 189)
65-69	148 (133, 164)	149 (133, 165)	198 (177, 219)
70-74	150 (135, 166)	199 (179, 219)	217 (194, 240)
75-79	122 (108, 136)	194 (179, 219)	223 (197, 249)
80-84	78 (67, 88)	140 (123, 158)	212 (183, 241)

LIVERPOOL  
UNIVERSITY  
LIBRARY

