OUTCOMES IN CLINICAL TRIALS IN CHILDREN WITH ASTHMA

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

by

Ian Sinha

April 2011
Abstract

Author: Ian Sinha

Thesis title: Outcomes in clinical trials in children with asthma

The selection of outcomes is a critically important decision when designing randomised controlled trials (RCTs). Informed clinical decisions can only be based on the results of RCTs that have measured outcomes of importance to both clinicians and patients. It can be difficult to know which outcomes should be measured in RCTs. Some groups advocate core outcome sets, which are a minimum set of outcomes that should be measured, and reported, in all clinical trials in a given condition. These increase the likelihood that important outcomes are measured, reduce non-uniformity between studies, and reduce the risk of outcome reporting bias.

We systematically reviewed studies that determined which outcomes to measure in clinical trials in children, and found that such work had been conducted in only few conditions, and the quality of existing work was variable. Few studies used structured consensus techniques to reach agreement about which outcomes to measure in trials, and parents were seldom involved. No studies included children. One condition in which there were no robust recommendations about which outcomes to measure
in RCTs was childhood asthma, which is a condition of considerable global importance.

We subsequently aimed to assess whether the absence of a core outcome set for RCTs of children with asthma meant that certain outcome domains were measured less frequently than others, and whether there was non-uniformity between studies in terms of outcomes selected. We conducted a systematic review of RCTs of children with asthma, published between January 1988 and December 2007, and found that the included studies focussed on short-term disease activity, but quality of life, functional status, and long-term outcomes were infrequently measured. Certain outcomes were measured and reported in various ways. We recommended that a core outcome set should be developed for childhood asthma, using structured consensus techniques, such as the Delphi process.

In order to aid the development of such a core set, we first systematically reviewed studies that used the Delphi process to determine which outcomes to measure in clinical trials. We observed variations in the methodology used, identified potential sources of bias, and provided recommendations about how such studies could be conducted and reported.

In order to develop a core outcome set for childhood asthma, we used a Delphi process to ascertain the views of 46 clinicians, and around 100 parents and young people, about which outcomes are most important and
relevant from their perspective, when making shared decisions about regular therapies which control asthma. The most important outcomes were symptoms, exacerbations, and quality of life. Although consensus still needs to be reached amongst other groups of individuals involved in clinical trials, we conclude that these outcomes should be measured, and reported, in all RCTs that aim to evaluate the effectiveness of regular therapies for children and young people with asthma.
Acknowledgements

I would like to start by sincerely thanking my supervisors, Professors Paula Williamson and Rosalind Smyth, for their continued support and guidance.

Several people have helped immensely with the work contained within this thesis. I thank Mrs Leanne Jones and Dr Ruairi Gallagher for working so hard as my co-reviewers, in Chapters 2 and 5 respectively. I also thank Miss Natalie Yates, Dr Bridget Young, and Professors Tony Marson, Doug Altman, and Mike Clarke, for their helpful comments on specific chapters. For help and support during the research described in Chapter 5, I thank Mrs Sadie Clayton, the BPRS (especially Professor Warren Lenney, Dr Iolo Doull, and Dr Jeremy Hull), Mrs Jenny Newman, Asthma UK, in particular Malayka Rahman, and Dr Jon Couriel, Dr David Heaf, Dr Kevin Southern, Dr David Lacy, Mrs Becky Bryson, and Mrs Elaine Kelly.

For my funding, I thank the NIHR MCRN Clinical Trials Unit and Coordinating Centre. I thank the British Lung Foundation for a travel grant, which enabled me to present at the American Thoracic Society Annual Scientific Meeting.

Finally, I must thank my parents, my wife Steph, and my daughter Sophie, for their constant love and support. I dedicate this thesis to my brother.
# Table of contents

ABSTRACT .................................................................................................................. I
ACKNOWLEDGEMENTS .............................................................................................. IV
TABLE OF CONTENTS ................................................................................................. V
LIST OF TABLES ........................................................................................................ VIII
TABLE OF FIGURES .................................................................................................. IX
ABBREVIATIONS USED IN THIS THESIS .............................................................. X

PUBLICATIONS AND PRESENTATIONS ARISING FROM THE WORK IN
THIS THESIS ........................................................................................................... XII

INTRODUCTION ........................................................................................................ 1
  1.1 OUTCOMES IN CLINICAL TRIALS ................................................................. 1
  1.2 OUTCOME DOMAINS ...................................................................................... 4
  1.3 DIFFERENT TYPES OF OUTCOMES .............................................................. 9
  1.4 STANDARDISING OUTCOMES FOR CLINICAL TRIALS ............................ 13
      1.4.1 Problems with heterogeneous selection, measurement and reporting
           of outcomes .............................................................................................. 13
      1.4.2 Standardising outcomes in trials ........................................................... 17
      1.4.3 Methods for reaching consensus ........................................................... 19
  1.5 DIFFICULTIES WITH SELECTING AND MEASURING OUTCOMES IN CLINICAL TRIALS IN CHILDREN ......................................................... 23
  1.6 ASTHMA IN CHILDREN .................................................................................. 26
      1.6.1 The impact of childhood asthma ............................................................ 26
      1.6.2 Clinical trials of regular therapies for children with asthma ................. 28
      1.6.3 The need for research to identify the most appropriate outcomes for
           clinical trials in children with asthma ...................................................... 33
  1.7 STRUCTURE OF THIS THESIS ....................................................................... 35

A SYSTEMATIC REVIEW OF STUDIES THAT AIM TO DETERMINE WHICH
OUTCOMES TO MEASURE IN CLINICAL TRIALS IN CHILDREN ..................... 37
  2.1 BACKGROUND ................................................................................................. 37
  2.2 AIMS ............................................................................................................... 39
  2.3 METHODS OF THE SYSTEMATIC REVIEW .................................................. 39
      2.3.1 Inclusion and exclusion criteria ............................................................. 39
      2.3.2 Identification of relevant studies ............................................................ 40
      2.3.3 Designing an appropriate search strategy .............................................. 42
      2.3.4 Selecting studies for inclusion in the review ......................................... 43
      2.3.5 Data Extraction ...................................................................................... 43
      2.3.6 Assessing the quality of included studies ............................................. 44
      2.3.7 Data Analysis ......................................................................................... 44
  2.4 RESULTS ......................................................................................................... 45
      2.4.1 Study selection ....................................................................................... 45
      2.4.2 Agreement between reviewers ............................................................. 47
      2.4.3 Description of included studies ............................................................. 48
A SYSTEMATIC REVIEW TO IDENTIFY WHICH OUTCOMES ARE MEASURED IN RANDOMISED CONTROLLED TRIALS IN CHILDREN WITH ASTHMA ................................................................. 69

3.1 BACKGROUND ............................................................................................................ 69
3.2 AIMS ............................................................................................................................ 70
3.3 METHODS OF THE REVIEW ....................................................................................... 71
  3.3.1 Inclusion and exclusion criteria .............................................................................. 71
  3.3.2 Identification of studies ......................................................................................... 71
  3.3.3 Search strategy ..................................................................................................... 72
  3.3.4 Data extraction and quality assessment .................................................................. 72
  3.3.5 Data analysis and presentation .............................................................................. 74
3.4 RESULTS ....................................................................................................................... 74
  3.4.1 Results of the search .............................................................................................. 74
  3.4.2 Description of included studies ............................................................................ 76
  3.4.3 Outcome domains measured in the studies .......................................................... 78
  3.4.4 Outcomes that were measured within the domains .............................................. 78
  3.4.5 Primary outcomes ............................................................................................... 82
  3.4.6 Outcomes measured in studies funded by the pharmaceutical industry ............... 83
  3.4.7 Trends over time in the selection of outcome domains ....................................... 83
3.5 DISCUSSION ................................................................................................................. 85
3.6 SUMMARY ..................................................................................................................... 89

A SYSTEMATIC REVIEW OF STUDIES THAT HAVE USED THE DELPHI PROCESS AS A METHOD FOR DETERMINING WHICH OUTCOMES TO MEASURE IN CLINICAL TRIALS ...................................................................... 91

4.1 BACKGROUND .............................................................................................................. 91
4.2 AIMS ............................................................................................................................ 93
4.3 METHODS .................................................................................................................... 93
  4.3.1 Inclusion and exclusion criteria ............................................................................ 93
  4.3.2 Identification of studies, and the search strategy .................................................. 94
  4.3.3 Data extraction .................................................................................................... 95
  4.3.4 Assessment of reporting quality .......................................................................... 95
  4.3.5 Data analysis and presentation of results ............................................................. 97
4.4 RESULTS ....................................................................................................................... 97
  4.4.1 Identification of studies ....................................................................................... 97
  4.4.2 Reporting quality ............................................................................................... 100
  4.4.3 Methodological variation between the studies .................................................... 104
4.5 DISCUSSION ............................................................................................................... 118
4.6 SUMMARY .................................................................................................................. 124
IDENTIFICATION OF THE MOST RELEVANT OUTCOMES FOR
EVALUATING REGULAR THERAPIES FOR CHILDHOOD ASTHMA .......... 126

5.1 BACKGROUND ........................................................................................................ 126
5.2 AIM .......................................................................................................................... 127
5.3 METHODOLOGICAL APPROACH ............................................................................ 127
5.4 PARTICIPANTS AND METHODS ............................................................................. 129
  5.4.1 Structure of the study ......................................................................................... 129
  5.4.2 Phase 1 .............................................................................................................. 130
  5.4.3 Phase 2 .............................................................................................................. 138
  5.4.4 Ethical considerations ......................................................................................... 141
5.5 RESULTS .................................................................................................................. 143
  5.5.1 Outcomes suggested in Phase 1 .......................................................................... 143
  5.5.2 Agreement between reviewers assessing the Phase 1 questionnaires ............... 149
  5.5.3 Phase 2: Ranking of outcomes in Phase 2 ......................................................... 159
5.6 DISCUSSION ............................................................................................................. 162
  5.6.1 Main findings ...................................................................................................... 162
  5.6.2 Comparison with the ATS/ERS outcomes taskforce ........................................... 164
  5.6.3 Robustness of the study ..................................................................................... 169
5.7 SUMMARY ............................................................................................................... 173

CONCLUSIONS AND FUTURE DIRECTIONS .............................................................. 175

6.1 MAIN CONCLUSIONS ............................................................................................... 175
6.2 IDENTIFYING THE BEST METHODS FOR DESIGNING CORE OUTCOME SETS ....... 177
6.3 DISSEMINATION AND IMPLEMENTATION OF CORE OUTCOME SETS: THE COMET
    INITIATIVE ................................................................................................................... 180
6.4 A CORE OUTCOME SET FOR PAEDIATRIC ASTHMA ........................................... 182
6.5 FINAL SUMMARY .................................................................................................... 189

LIST OF REFERENCES .................................................................................................. 190

APPENDICES .................................................................................................................. 218

APPENDIX 1 – SEARCH STRATEGIES USED IN THE SYSTEMATIC REVIEW OF STUDIES
    THAT AIMED TO DETERMINE WHICH OUTCOMES TO MEASURE IN CLINICAL TRIALS IN
    CHILDREN (CHAPTER 2) ............................................................................................. 219
APPENDIX 2 – STUDIES EXCLUDED FROM THE REVIEW DESCRIBED IN CHAPTER 2 .... 223
APPENDIX 3 - DESCRIPTION OF STUDIES INCLUDED IN THE SYSTEMATIC REVIEW
    DESCRIBED IN CHAPTER 2 ......................................................................................... 230
APPENDIX 4 – OUTCOMES SELECTED IN THE STUDIES INCLUDED IN THE SYSTEMATIC
    REVIEW DESCRIBED IN CHAPTER 2 .......................................................................... 239
APPENDIX 5 – SEARCH STRATEGY USED IN THE SYSTEMATIC REVIEW, DESCRIBED IN
    CHAPTER 3, OF RCTs OF INHALED CORTICOSTEROIDS FOR CHILDREN WITH ASTHMA
    ..................................................................................................................................... 245
APPENDIX 6 – PHASE 1 QUESTIONNAIRE, AND INVITATION, SENT TO CLINICIANS ....... 246
APPENDIX 7 - PHASE 1 QUESTIONNAIRE DISTRIBUTED TO PARENTS ......................... 250
APPENDIX 8 – QUESTIONNAIRES DISTRIBUTED TO CLINICIANS AND PARENTS IN PHASE
    2 ................................................................................................................................. 254
APPENDIX 9 – SUMMARY OF RESULTS FROM ASTHMA UK ..................................... 262
APPENDIX 10 – CORRESPONDENCE WITH NATIONAL RESEARCH ETHICS SERVICE .... 263
APPENDIX 11 – COPY OF PUBLICATIONS ARISING FROM THE WORK IN THIS THESIS .... 265

VII
List of tables

<table>
<thead>
<tr>
<th>Number</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1 List of studies included in the review</td>
<td>49</td>
</tr>
<tr>
<td>Table 2 Quality of reporting in studies that aim to determine which outcomes to measure in clinical trials in children</td>
<td>58</td>
</tr>
<tr>
<td>Table 3 Characteristics of included studies</td>
<td>77</td>
</tr>
<tr>
<td>Table 4 Frequency with which outcomes were reported in the 115 RCTs which involved only children</td>
<td>80</td>
</tr>
<tr>
<td>Table 5 Summary of the reporting quality of the studies which used the Delphi technique to determine which outcomes or domains to measure in clinical trials or systematic reviews</td>
<td>101</td>
</tr>
<tr>
<td>Table 6 Size and composition of the groups involved in Delphi processes to determine which outcomes to measure in clinical trials</td>
<td>106</td>
</tr>
<tr>
<td>Table 7 Level of anonymity in the studies, and whether participants met at any point</td>
<td>109</td>
</tr>
<tr>
<td>Table 8 How consensus was reached about which outcomes to measure in clinical trials</td>
<td>114</td>
</tr>
<tr>
<td>Table 9 Outcomes suggested in Phase 1 by clinicians, for pre-school children</td>
<td>144</td>
</tr>
<tr>
<td>Table 10 Outcomes suggested in Phase 1 by clinicians for school-aged children</td>
<td>145</td>
</tr>
<tr>
<td>Table 11 Frequency with which outcomes were suggested by parents of pre-school children (n=11)</td>
<td>146</td>
</tr>
<tr>
<td>Table 12 Frequency with which outcomes were suggested by parents of school-aged children (n=27)</td>
<td>147</td>
</tr>
<tr>
<td>Table 13 Frequency with which outcomes were suggested by young people (n=11)</td>
<td>148</td>
</tr>
<tr>
<td>Table 14 Whether outcomes were carried forward to Phase 2: determined by the frequencies with which outcomes were suggested by participants at the end of Phase 1, and with which they were measured in RCTs of inhaled corticosteroids for children with asthma</td>
<td>156</td>
</tr>
<tr>
<td>Table 15 Preschool children: the importance of each outcome listed in the Phase 2 questionnaire, as scored by clinicians and parents</td>
<td>160</td>
</tr>
<tr>
<td>Table 16 School-age children: The importance of each outcome listed in the Phase 2 questionnaire, as scored by clinicians and parents</td>
<td>161</td>
</tr>
<tr>
<td>Table 17 Outcomes recommended by the ATS/ERS, categorised by whether they should be considered essential, desirable or optional</td>
<td>166</td>
</tr>
<tr>
<td>Table 18 Comparison between the top 6 outcomes for school-aged children, identified after Phase 2 of our study, and the recommendations of the ATS/ERS taskforce</td>
<td>167</td>
</tr>
</tbody>
</table>
## Table of figures

<table>
<thead>
<tr>
<th>Number</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Flowchart of the review: identification of studies that aim to determine which outcomes to measure in clinical trials in children</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Flowchart of the review: identification of RCTs, assessing ICS in children with asthma, published between January 1988 and December 2007</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Trends in the measurement of outcome domains between 1988 and 2007</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Flowchart of the review: Identification of studies that used the Delphi process to determine which outcomes or domains to measure in clinical trials or systematic reviews</td>
</tr>
<tr>
<td>Figure 5</td>
<td>Flowchart showing participants in Phases 1 and 2</td>
</tr>
</tbody>
</table>
Abbreviations used in this thesis

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AHCH</td>
<td>Alder Hey Children’s Hospital</td>
</tr>
<tr>
<td>APH</td>
<td>Arrowe Park Hospital</td>
</tr>
<tr>
<td>ATS</td>
<td>American Thoracic Society</td>
</tr>
<tr>
<td>BDP</td>
<td>Beclomethasone Diproprionate</td>
</tr>
<tr>
<td>BPRS</td>
<td>British Paediatric Respiratory Society</td>
</tr>
<tr>
<td>COMET</td>
<td>Core Outcome Measures in Effectiveness Trials</td>
</tr>
<tr>
<td>CSG</td>
<td>Clinical Study Group</td>
</tr>
<tr>
<td>DSCG</td>
<td>Disodium Cromoglycate</td>
</tr>
<tr>
<td>EMA</td>
<td>European Medicines Agency</td>
</tr>
<tr>
<td>ERS</td>
<td>European respiratory Society</td>
</tr>
<tr>
<td>ENO</td>
<td>Exhaled Nitric Oxide</td>
</tr>
<tr>
<td>ERS</td>
<td>European Respiratory Society</td>
</tr>
<tr>
<td>FDA</td>
<td>Food and Drug Administration</td>
</tr>
<tr>
<td>FEV1</td>
<td>Forced Expiratory Volume in one second</td>
</tr>
<tr>
<td>ICS</td>
<td>Inhaled corticosteroid</td>
</tr>
<tr>
<td>JLA</td>
<td>James Lind Alliance</td>
</tr>
<tr>
<td>LABA</td>
<td>Long acting Beta₂ Agonist</td>
</tr>
<tr>
<td>LTRA</td>
<td>Leukotriene Receptor Antagonist</td>
</tr>
<tr>
<td>MCRN</td>
<td>Medicines for Children Research Network</td>
</tr>
<tr>
<td>NDDI</td>
<td>Neonatal Drug Development Initiative</td>
</tr>
<tr>
<td>NGT</td>
<td>Nominal Group Technique</td>
</tr>
<tr>
<td>NICE</td>
<td>National Institute for Health and Clinical Excellence</td>
</tr>
<tr>
<td>OMERACT</td>
<td>Outcome Measures in Rheumatology</td>
</tr>
<tr>
<td>ORB</td>
<td>Outcome Reporting Bias</td>
</tr>
<tr>
<td>PEFR</td>
<td>Peak Expiratory Flow Rate</td>
</tr>
<tr>
<td>PIP</td>
<td>Paediatric Investigation Plan</td>
</tr>
</tbody>
</table>
QoL  Quality of Life  
RCT  Randomised Controlled Trial  
WHO  World Health Organization  

Initials of researchers involved in the work in this thesis:  
IS  Ian Sinha  
LJ  Leanne Jones  
RG  Ruairi Gallagher  
RLS  Rosalind Smyth  
PRW  Paula Williamson
Publications and presentations arising from the work in this thesis

The work contained in Chapter 2 has been published in PLoS Medicine (Sinha et al 2008), and presented at the Royal College of Paediatrics and Child Health Annual Scientific meeting (York, 2008), the James Lind Alliance conference “Outcomes in Clinical Trials: whose responsibility?” (London, 2008), and the Cochrane Colloquium (Freiburg, 2008).

The work contained in Chapter 3 has been published in PLoS One (Sinha et al 2009) and has been presented at the American Thoracic Society Annual Scientific Meeting (San Diego 2009).

The work contained in Chapter 4 has been published in PLoS Medicine (Sinha et al 2011).

The work contained in Chapter 5 has been submitted for publication, and was presented at the launch meeting of the COMET initiative (Liverpool 2010, the Royal College of Paediatrics and Child Health Annual Scientific Meeting (Warwick, 2010), and the Cochrane Colloquium (Denver, 2010).

A copy of the publications arising from the work in Chapters 2, 3, and 4 is included in Appendix 12.
Chapter 1

Introduction

1.1 Outcomes in clinical trials

Interventions used in healthcare may have various beneficial and harmful effects. These are known, in clinical practice and research, as outcomes. Interventions in randomised controlled trials (RCTs) are evaluated in terms of the effect they have on pre-determined outcomes. It is critically important, therefore, that the right outcomes are measured in RCTs.

Primary and secondary outcomes

The primary outcome is an integral component of the research question. It should be explicitly stated and described, because it reflects the objective of the trial, and determines the sample size required (Moher et al 2001). There is usually only one primary outcome, which should relate to the research question. In some situations, more than one primary outcome can be measured, in order to assess treatments more comprehensively. This can be useful if it unclear which single primary outcome will best answer the research question. For example, in trials of patients with schizophrenia, one co-primary outcome can measure a patient’s cognitive ability, while the other assesses functional ability (Green et al 2008). In other situations, one
co-primary outcome may measure the efficacy of an intervention, and the other its safety, such as in trials of therapies for children with familial hypercholesterolemia (Tonstad et al 1996).

Most trials also measure secondary outcomes, in order to evaluate other beneficial or harmful effects of treatments. These must be clearly stated and described in the study protocol and report (International Conference on Harmonization 1999). They should be used judiciously, because the more that are measured in an RCT, the more likely it becomes that one will show a statistically significant result because of chance alone (Pocock 1997).

**Why RCTs must measure important outcomes in order to inform clinical decisions**

Justifying the use of a healthcare intervention depends on whether its beneficial effects are important, and sufficiently outweigh its risks. As RCTs are the most valid scientific method for evaluating these effects, their results are used to inform such decisions (Guyatt, & Rennie 2002). In order for the results of RCTs to be useful, they should measure outcomes that are relevant to people involved in making clinical decisions about an individual patient, group, or population.

For individual patients, clinicians should integrate the best available research evidence to determine which interventions are most helpful and safe (Sackett et al 1996). This evidence should be derived from well
designed, rigorously conducted RCTs, which will best inform practice if they have measured outcomes that patients and clinicians find important. To decide whether or not an outcome is important to patients, the following has been suggested (Guyatt, & Meade 1997):

“Imagine suggesting to patients that the outcome in question, and no other, would change with treatment. If patients would be willing to undergo whatever risks and inconvenience are associated with the treatment, and willing to pay ... the associated costs, then the outcome is important to them”.

This definition is even more pertinent today, because the medical profession has observed a shift from the traditional ‘paternalistic’ relationship, in which clinicians decide what is best for patients, towards a situation in which they make shared decisions about management. This change was initiated on ethical grounds, as it was felt that patients had a right to be involved in decisions about their own health (Charles et al 1997; Frosch, & Kaplan 1999). Evidence now suggests that involving patients in management decisions also improves their satisfaction with their care, adherence to treatment, and overall health (Joosten et al 2008; Hamann et al 2003). In order for patients to share in making informed decisions about an intervention, they should know whether it is likely to benefit them in the way that they would hope, and this relies on the measurement of outcomes, in RCTs, which they feel are important.
RCTs are also used to inform decisions about whether interventions are appropriate for populations of patients. In the UK, the National Institute for Health and Clinical Excellence (NICE) design clinical guidelines, and approve or reject funding for interventions, based on whether their benefits are worth their financial cost (Claxton et al 2002; Drummond 2002). Drug regulatory authorities, such as the Food and Drug Administration (FDA), and European Medicines Agency (EMA), will grant a marketing authorisation for a product to be used in their area of jurisdiction if its benefits outweigh its risks. In order to enable these agencies to make such assessments accurately, RCTs should provide information about important beneficial and harmful treatment effects that are likely to affect the patient’s health or overall well-being.

1.2 Outcome domains

Outcome domains are frameworks, which can be used to classify broad aspects of the effects of interventions. It may not be clear which domains should be measured in an RCT, and several outcomes may be relevant to a single domain. The ways in which clinicians and researchers consider the effects of illnesses have changed over time, and this has led to the evolution of certain important outcome domains that are measured in RCTs.
Outcomes reflecting disease processes: the biomedical model

For much of the 20th Century, the effects of illnesses were considered in a biological and physical framework, now known as the biomedical model. The basis for this model originated in the 19th Century, when medical advances, notably in the field of pathology, enabled clinicians and researchers to consider illnesses in a more scientific manner (Wade, & Halligan 2004). This is reflected by the outcomes measured in the earliest controlled clinical trials in the modern era, which were conducted by the Medical Research Council in the 1940s (Chalmers, & Clarke 2004).

In the first of these, patients suffering from common cold were randomised, by alternate allocation, to receive either Patulin or placebo solution for two days. After one, two, and seven days, they were assessed to see if their symptoms had resolved (Medical Research Council 1944). The second trial evaluated whether treatment with Streptomycin and bed-rest for six months, when compared with six months of bed-rest alone, effectively treated pulmonary tuberculosis (Medical Research Council 1948). Radiological findings, clinical condition, temperature, weight, sedimentation rate, and sputum culture, were compared in survivors after six months. The main analysis was based on radiological changes, as the authors suggest that this was “the single most important factor to consider”.

Outcomes measuring biological effects of interventions remain important today. All four RCTs of medical treatments for pulmonary tuberculosis, published in 2009, measured clearance of bacteria from sputum, as the

The effects of illness on functional status: the biopsychosocial model

The effects of interventions on a patient’s ability to perform normal daily functions has increasingly been recognised as an important domain in clinical trials (Duncan et al 2000; Gandhi et al 2008).

The earliest studies of the non-biological aspects of illness examined the levels of assistance required by elderly people (New York Dept of Social Welfare 1937), and the functional ability of patients with heart disease (New York Heart Association 1939). The publication of a study of nitrogen mustard, as a palliative treatment for patients with cancer (Karnofsky et al 1948), is probably the earliest report of the evaluation of the effects of an intervention on functional status (Prutkin, & Feinstein 2002).

In the 1960s, psychologists and clinicians felt that the biomedical model did not reflect the complex relationship between biological, social and emotional aspects of illness, and this led to the development of the biopsychosocial model of medicine. Also in the 1970s, the World Health Organization (WHO) recognised the need to change how they assessed the health of populations, because its member states, expected to routinely compile health statistics, reported only mortality rates from various diseases. The publication of the International Classification of Impairments, Disabilities and Handicaps (WHO 1980) represented a
formal move to consider effects of disease not only in terms of anatomical and physiological abnormalities ("impairments"), but also the resulting restriction of activities ("disability"), and its consequences for an individual’s ability to function, in society, as they would like ("handicap").

A landmark study, which emphasised the importance of measuring functional status in clinical research, was The Health Insurance Experiment (Brook et al 1983). This large RCT, conducted between 1971 and 1982, compared medical insurance plans in the USA. Researchers found that, by measuring patients’ perceptions of their health and functional status, they generated useful information, which would not have been identified from traditional biological endpoints. The subsequent Medical Outcomes Study, which evaluated 20,000 patients’ own reports of their health and functional status, also recommended that the primary goal of treating chronic conditions should be to “maximise function in everyday life and to achieve the highest level of well-being” (Tarlov et al 1989; Stewart et al 1989).

The overall well being of patients: the psycho-sociological model

The first use of the term “quality of life” (QoL) in the medical literature was in an article relating to patients receiving haemodialysis, in which the author comments that “for most, the quality of life was unacceptable” (Retan, & Lewis 1966). In the same year, an RCT, in which women with breast cancer were treated either with radical mastectomy, or more conservative surgery, was published. One secondary outcome was the
“quality of survival”, as measured by their feelings about their health and their treatment (Eisenberg 1966). This was probably the first time emotional well-being was measured as an outcome in a clinical trial (Prutkin, & Feinstein 2002). Until the 1980s, QoL slowly became a more prominent outcome in medical research. Medical researchers measured emotional well-being scales developed by psychologists, but physical and functional consequences of illness on QoL were infrequently studied in conjunction with these (Prutkin, & Feinstein 2002).

Two events in the 1980s catalysed the development of QoL as a clinical trial outcome (Prutkin, & Feinstein 2002; Willke et al 2004). Firstly, the FDA announced that QoL should be measured as a “key efficacy parameter” in clinical trials of anticancer agents, which had traditionally focussed on length of survival. Furthermore, the FDA stated that if a therapy was shown to improve QoL, even if it did not have increase survival, it could still be granted a marketing license (Temple 1995; Johnson, & Temple 1985). Secondly, the first RCT to measure QoL, as a primary outcome, was published (Croog et al 1986). The effects of three antihypertensive agents on QoL were compared, using validated measures of satisfaction with life, physical state and social functioning. When patients who were randomised to receive one drug had significantly better QoL than those in other groups, the pharmaceutical industry realised that products could be marketed for their effects on QoL, as well as their physiological outcomes, and clinical researchers learnt that QoL could be measured, in a meaningful and valid way, in clinical trials. Since the 1980s, the number of RCTs measuring QoL has increased, but there are still
concerns that it is measured infrequently, using non-validated tools (Clarke, & Eiser 2004; Sanders et al 1998).

1.3 Different types of outcomes

As well as determining which domains to measure in trials, researchers must also decide on the most appropriate outcomes with which they should be measured. This can be difficult, because a variety of outcomes could be measured within each domain. As well as the importance of an outcome, trialists must consider whether it is relevant and acceptable for the patients included in the trial, responsive to the interventions being compared, and appropriate for the trial itself, in terms of the length of follow-up required, the financial cost, and whether validated methods are available for measuring it. The decision is made more complex by the various types of outcomes from which trialists can choose, which are described below.

Clinical outcomes and surrogates

A clinical outcome relates to how a patient feels, functions, or survives. A surrogate outcome is a laboratory measurement or physical sign that is used as a substitute for a clinical outcome (Atkinson et al 2001; Temple 1995). A biomarker is a laboratory measurement that is used to predict a biological, pathogenic, or physiological outcome, rather than a clinical one (Lassere et al 2007). Trialists must decide which of these types of
outcomes is most appropriate for both the research question and the specific clinical trial.

In late phase clinical trials, the primary outcome should reflect the effects of an intervention on an important clinical outcome. If the clinical outcome occurs a long time after the intervention is started, however, this may be difficult to achieve. Surrogate outcomes, which are affected by interventions earlier than clinical ones, can be used to predict these effects. This means that fewer patients need to be recruited, and less time is required, to conduct a clinical trial. As well as making clinical trials cheaper and easier to conduct, this can lead to faster approval of interventions by drug regulatory agencies. In certain clinical areas, in which effective therapies are lacking, such as oncology, there may be an urgency to make interventions available to patients quickly. In other areas, such as trials of anti-hypertensive agents for adults, the ultimate aim is to identify which treatments reduce the risk of mortality. In these situations, surrogate outcomes and biomarkers are used to accelerate the drug development process, as trials may need to be conducted for several decades to determine such effects.

Despite these theoretical advantages, surrogate outcomes have been the subject of much debate since the mid-1990s. When they have been criticised, the main limitation has been that their correlation with the clinical outcome is inaccurate (Holloway, & Dick 2002; Fleming, & DeMets 1996; D'Agostino 2000), or affected by confounding factors (Baker, & Kramer 2003).
**Measured and reported outcomes**

Outcomes can be categorised by the kind of treatment effect they measure. Some are objective, biological outcomes, such as physiological measures of lung function in patients with asthma, or laboratory measures of glycosylated haemoglobin in patients with diabetes. Others, known as clinician reported outcomes, may be interpreted or observed by a medical caregiver. For example, a clinician could observe a patient (eg resolution of an infection), interpret an investigation (eg radiological results), or interpret information given by patients in order to complete a scale or score (eg the global physician response used in RCTs in juvenile arthritis (Giannini et al 1997)). Patient reported outcomes measure how a patient feels or functions, as described directly by the patient, without interpretation by a clinician or investigator. Commonly, these are measures of the patient’s view of their symptoms, functional status, or quality of life (Willke et al 2004).

Although laboratory and physiological measurements may be more objective, easier, and cheaper to measure, they usually provide a less comprehensive view of the effects of an illness than outcomes reported by clinicians or patients. Patient reported outcomes are also important because a patient’s view of how an intervention has affected their illness, or improved their life, may differ from their clinician’s opinion.
Composite outcomes

Composite outcomes comprise more than one endpoint, outcome or domain. One advantage of this approach is that composite outcomes may enable trialists to combine clinical events into a single primary outcome, and report the number of patients experiencing any one of these, so a smaller sample size is required for the trial. Another advantage is that a composite outcome may provide a more comprehensive evaluation of the treatment itself (Montori et al 2005). This may be by combining separate events, such as serious complications of prematurity (Anand et al 2004) or, as in the case of QoL, by capturing various continuous outcomes.

Despite these advantages, composite primary outcomes can lead to misleading results, usually when the effects of the treatment on constituent components of the outcome differ. This may arise if the overall composite outcome improves, but the intervention does not affect the most important component, which is usually mortality. For this reason, each constituent component should be clinically meaningful (Cannon 1997), and should be analysed and reported separately as a secondary outcome (International Conference on Harmonization 1999; Freemantle et al 2003).
1.4 Standardising outcomes for clinical trials

The selection and measurement of outcomes may vary across clinical trials in the same condition, and this can cause a variety of problems, which are discussed in this section.

1.4.1 Problems with heterogeneous selection, measurement and reporting of outcomes

A variety of domains and outcomes can be selected for clinical trials, and these can be measured, analysed, and reported in different ways. Three main problems can arise if the selection, measurement and reporting of outcomes is non-uniform across clinical trials.

Important outcomes can be overlooked

If trialists are not required to adhere to an accepted list of mandatory outcomes, it is likely that factors relating to the conduct of the trial, such as the sample size, will determine which outcomes are measured. The relative importance of outcomes then becomes a secondary consideration. As a result, outcomes that are favourable for researchers may be selected, but those that are important to patients or clinicians could be overlooked.
In a systematic review to evaluate the methodological quality of 199 RCTs in diabetes, published between 1987 and 2003, the authors categorised the primary and secondary outcomes as patient-important (defined as those expected to directly affect QoL), physiological measures that did not predict patient-important outcomes, and surrogate outcomes (Montori et al 2006). With regard to the primary outcomes, only 42/199 (21%) studies measured patient important outcomes, 118/199 (59%) measured physiological outcomes and 38/199 (19%) measured surrogate outcomes. The same group later evaluated outcomes in 436 ongoing trials in diabetes, and found that 78/436 (18%) measured patient important outcomes, 69/436 (16%) measured physiological outcomes, and 268/436 (61%) measured surrogate outcomes (Gandhi et al 2008).

Another systematic review assessed the outcomes measured in 174 clinical trials of acute therapies for stroke, published between 1955 and 1995 (Roberts, & Counsell 1998). Mortality was reported in 132/174 (76%), and outcomes reflecting physical impairment were reported in 133/174 (76%), but measures of disability were reported in 73/174 (42%) and measures of handicap or QoL in just 4/174 (2%). The author of another systematic review, of 51 RCTs of acute therapies for stroke, identified similar problems (Duncan et al 2000).

The authors of two of these systematic reviews recommended that a potential solution to these problems would be to develop core set of outcomes, incorporating outcomes of direct importance to patients, which
should be measured and reported in all clinical trials of diabetes (Gandhi et al 2008) and acute stroke (Duncan 2000).

**Impaired meta-analysis**

When synthesising results of RCTs, in meta-analyses, problems arise if included studies measure different outcomes from each other. For example, in a Cochrane review of trials evaluating therapies for acute stroke, the primary outcome of the review was only reported in study reports of 9/22 of the included studies (Gubitz et al 2000). This represents a missed opportunity in the other 13 trials to contribute data to an important research question.

Even if the same outcomes are selected, they may be measured and analysed in different ways, and this too can impair the ability to synthesise their results. One descriptive review, of 2000 trials of interventions for patients with schizophrenia, observed that 640 different scoring scales had been used (Thornley, & Adams 1998). In a subsequent Cochrane review, the same authors recommended that “a concerted effort should be made to agree on which measures are the most useful. Studies within this review reported on so many scales that, even if results had not been poorly reported, they would have been difficult to synthesise in a clinically meaningful way” (Adams et al 2007).

A Cochrane review of probiotics for the prevention of antibiotic-associated diarrhoea in children included studies that varied considerably in the
definition of diarrhoea itself. As a result, one secondary outcome of the review included 4/10 studies, and the other included 3/10 (Johnston et al 2007). For similar reasons, authors have recommended the development of core outcome sets in acute asthma (Blitz et al 2005) and ulcerative colitis (Cooney et al 2007).

**Selective reporting of results**

In the absence of a set of core outcomes, which must be reported in all clinical trials in a given condition, trialists may decide to omit certain results from the final study report. The practice of selective reporting of a subset of the original recorded outcomes, based on the results, is called outcome reporting bias (ORB) (Hutton, & Williamson 2000).

A systematic review, of studies that examined ORB in published RCTs, found that statistically significant results are more likely to be fully reported than non-significant ones, and it appears to be common practice to change or omit the original primary outcome, or introduce a new one, as the study progresses (Dwan et al 2008).

Further problems from ORB arise when systematic reviewers attempt to meta-analyse the results of RCTs. When conducting Cochrane reviews, authors are now required to assess the risk of ORB in included studies. A recent study examined the impact that ORB might have on the analysis of the primary outcome in 283 Cochrane reviews, published in 2006 and 2007 (Kirkham et al 2010). One third of reviewers had suspected ORB in at
least one of the reports of studies included in the primary analysis of the review. Furthermore, of the meta-analyses in which the review primary outcome showed statistically significant result, 8/42 (19%) became non-significant after adjustment for outcome reporting bias, and 11/42 (26%) would have significantly overestimated the treatment effect, by at least 20%. The authors recommended that core outcome sets could prevent selective reporting, and make it easier to assess the risk of ORB in trial reports, by informing the reader which outcomes should have been measured, and hence which results should be reported.

1.4.2 Standardising outcomes in trials

The problems associated with non-uniform measurement, selection, and reporting of outcomes can be prevented by standardising outcomes across clinical trials in a given condition. Two notable examples of groups that have attempted to achieve this are described below.

WHO outcomes for cancer

Probably the earliest initiative to standardise outcomes for trials was in oncology. The WHO recommended that the ability to compare results between trials would be improved if trialists measured and reported a minimum core set of outcomes. International meetings were held in 1977 (Turin) and 1979 (Brussels), involving 36 clinicians and researchers, with
the aim of reaching consensus about which outcomes should be measured, and how they should be reported (Miller et al 1981). The panel recommended that, as a minimum, trialists must report the measured response of the tumour and metastases (including whether these completely or partially resolved), duration of response to treatment, acute and long-term adverse effects of therapy, and a measure of the patient’s emotional well-being.

OMERACT

Probably the most notable examples of work to develop core outcome sets have been conducted in rheumatology. In 1992, rheumatologists from Europe and North America observed that RCTs in rheumatoid arthritis varied in the selection and measurement of outcomes, largely dependant on which continent they were conducted in, and that this impaired comparison and meta-analysis of results. This led to a conference, in 1992, at which around 80 participants aimed to agree on the minimum outcomes that should be measured and reported in these trials (Tugwell, & Boers 1993). The collaboration was known as OMERACT (Outcome Measures in Rheumatoid Arthritis Clinical Trials). Since the initial conference, OMERACT has held 10 conferences at which core outcome sets are designed for various conditions in rheumatology.

The OMERACT core outcome sets are based on evidence about which outcomes are most appropriate. Interested individuals collate and
summarise existing literature about domains and outcomes, and then a wider group of clinicians and researchers reach consensus about which outcomes should be included in the core set, and how they should be measured. This consensus is reached using structured techniques, discussed in the next section, which are likely to improve the quality and validity of the recommendations.

1.4.3 Methods for reaching consensus

Ideally, guidelines and recommendations are based on evidence derived from rigorously conducted empirical studies. In situations where the best recommendations are uncertain, or such studies have not been conducted, decisions made by a group of people are likely to be more credible than those made by one individual, because they are based on a wider range of experience and knowledge (Murphy et al 1998).

The three main methods for reaching formal consensus in health research are open group discussions, Nominal Group Technique (NGT), and Delphi process. These are described below.

Open group discussions

There are three main variations of open group discussion. In the informal discussion group, interaction between participants is not structured (Murphy et al 1998). In the structured group, the discussion is steered, by a
facilitator, through a series of problem-solving steps (Murphy et al 1998; Jarboe 1996). In the NIH consensus development conference, a group, of around ten participants, discusses relevant evidence, before voting to reach agreement on a particular topic (Fink et al 1984; Murphy et al 1998).

The aim of open group discussion is to stimulate debate amongst participants. Types of personalities within the group, however, can inhibit the group’s productivity, and compromise the credibility of the pooled judgement. These may include participants who dominate the decision making process, are unwilling to change their position once publicly taken, do not voice disagreement with people more senior than themselves, acquiesce for the sake of reaching consensus, or are reluctant to mention a new idea, in case they are criticised (Linstone et al 1975; Thangaratinam, & Redman 2005).

**Nominal Group Technique (NGT)**

NGT is a method of structuring interaction within a group. It was first developed because unstructured face-to-face discussion methods can inhibit members from speaking freely, and groups may focus on one particular idea, at the expense of others, when discussing a problem (Murphy et al 1998; Delbecq, & Van de Ven 1971).

In the NGT, a group of individuals, typically between 10 and 15 in number, discuss one clearly defined problem at a time, during a face-to-face meeting. First, each participant shares their opinion with the group. After
this, each participant votes, anonymously, and the group opinion is compiled. There should be a pre-determined definition of what constitutes consensus, disagreement, or uncertainty amongst the group (Murphy et al 1998).

The advantage of NGT is that specific problems can be discussed, in order to generate ideas and share opinions, with each participant’s view incorporated equally in the vote. This means that the final group decision is less likely to be swayed by a vocal minority, or by people of seniority. The disadvantages are that only a small group can be involved in a single NGT, and they must all convene at the same place and time.

**Delphi technique**

The Delphi technique comprises a series of questionnaires, which are answered, anonymously, by a panel of participants. After each questionnaire, the group response is fed back to participants (Dalkey, & Helmer 1963).

With regard to the overall validity of the final consensus, this approach confers certain advantages over less structured methods of reaching consensus. Participants in a Delphi study do not interact directly with each other, so situations where the group is dominated by the views of certain individuals can be avoided. When participants consider whether to change their opinion, after the overall group response is fed back following each questionnaire, this decision is not affected by the perceived need to be seen
to agree with senior, overly vocal, or domineering individuals. As with NGT, every person’s opinion is reflected equally in the results.

The Delphi method was originally designed by the RAND corporation between the 1940s and 1960s (Dalkey, & Helmer 1963; Dalkey 1969). RAND is an American institute, formed in the aftermath of the Second World War, which facilitates research between government agencies, military forces, industries and universities (Campbell 2004). The Delphi method was initially designed as a forecasting tool for predicting Russian military strategies between the 1940s and 1960s, because it was thought that predictions made by a group of people were more likely to be correct than those made by the same individuals working in isolation. In the 1960s, Delphi methodology was mainly used for predicting social and technological change (Ament 1970), and after this the research areas for which it was used diversified (Linstone et al 1975; Needham, & Loë 1990; Rieger 1986; Mullen 2003).

In the mid 1970’s, the validity of the Delphi technique was widely scrutinised (Hill, & Fowles 1975; Coates 1975; Goldschmidt 1975; Sackman 1975). Critics suggested that the fundamental characteristics of the Delphi technique were flawed, because ‘expertise’ is an arbitrary concept, anonymity leads to lack of accountability, and hence poor quality answers, and participants are led towards conformity with the group, rather than consensus of true opinions. Another criticism was that people who used the Delphi technique as a research tool were not adhering to accepted principles of scientific investigation, and that approaches to aspects of the
methodology, especially population sampling techniques and psychometric validation of questionnaires, were inadequate.

Despite this period of scrutiny the Delphi technique continued to be widely used. It has been adapted to work well via electronic media, and improvements in global communication have made it feasible to involve geographically distant participants, in larger numbers than are traditionally used in open group discussions and NGT. It is increasingly being used to reach consensus around topics in medicine, such as education (Alahlafi, & Burge 2005), clinical guidelines (Morita et al 2005), and prioritisation of research topics (Kellum et al 2008).

1.5 Difficulties with selecting and measuring outcomes in clinical trials in children

Conducting clinical trials in children presents unique challenges (Caldwell et al 2004; Smyth 2001), including difficulties in the selection and measurement of outcomes. These can occur for various reasons.

One problem with the selection of outcomes for RCTs in children is that it may be influenced by the fact that it is often difficult to recruit patients. This is partly because the number of children suffering from chronic conditions is less than that in adults, so there is a smaller pool of patients from which to recruit (Smyth 2001). Problems may also arise because of reluctance from parents and clinicians to enrol children in research studies
(Caldwell et al 2004). As a consequence of the smaller sample size available, trialists may have to select a primary outcome based on feasibility, rather than importance.

One approach to the small numbers of children participating in clinical research is to include them in RCTs conducted primarily in adults, possibly analysing them as a subgroup. This approach may not be appropriate, however, for two main reasons.

Firstly, the outcomes and domains that are relevant and important for adults may not be suitable for children. Furthermore, the measurement of certain outcomes, such as physiological tests, or those requiring invasive investigation, may not be tolerated, feasible, or acceptable in children. Conversely, certain outcomes may be unique to a paediatric population, and may not be measured in a study primarily focussed on adults. For example, attainment of developmental milestones is an outcome that is unique to babies, infants, and pre-school toddlers, and growth to an optimal adult height is an outcome of importance in many chronic conditions in childhood.

Secondly, illnesses affect children and adults in different ways. This means that the same outcome domains may need to be measured in different ways. The factors affecting functional status, for example, will differ between adults and children, because their perceived roles in society and the family are different.
Even amongst children, there are several distinct age brackets, in which different outcomes for evaluating treatments may be relevant. Sepsis is one example of a condition in which the relevant outcomes may vary between age groups. In preterm neonates, infections are frequently fatal, and so mortality would be the most important outcome. In older children, mortality from sepsis is much less common, and so outcomes reflecting treatment effects such as need for intensive care, long-term complications, or QoL after survival, may be more relevant.

Whether an outcome is selected for an RCT in children may depend on whether or not there are age-appropriate, validated tools with which to measure it. The length of follow-up required to evaluate important long-term outcomes, manifest in later childhood or adulthood, may preclude their direct measurement. In these situations, surrogate outcomes are often measured in place of the long-term clinical outcome. These surrogate outcomes, however, may not have been robustly validated to assess whether they accurately predict the long-term clinical outcome. Similar problems, of whether or not outcomes and instruments have been validated for use in a paediatric population, may occur with patient reported outcome measures (PROMs).
1.6 Asthma in children

Much of the work in this thesis relates to outcomes in clinical trials in children with asthma. Section 1.6.1 discusses the impact of childhood asthma, and Section 1.6.2 discusses why it can be difficult to know which outcomes to measure in RCTs of pharmacological interventions for children with this condition.

1.6.1 The impact of childhood asthma

Asthma is a complex, chronic disorder of the airways. Inflammation and structural damage lead to airway obstruction and hyper-responsiveness. These cause frequent symptoms of cough, wheeze and shortness of breath, which are episodically worsened, acutely, by a variety of common triggers (Murray et al 2000; Forfar et al 2003). Asthma is the commonest chronic condition of childhood, and its prevalence has been increasing since the 1980s (Akinbami, & Schoendorf 2002). The annual direct and indirect financial costs of asthma were estimated, in the US, to be 5.1 billion dollars in 1997.

Symptoms of asthma are often frequent and troublesome. The cough, which is typically nocturnal, can disrupt sleep for the child and other family members. Shortness of breath and poor exercise tolerance affect the child’s ability to play sport, partake in normal activities, or adhere to as
healthy a lifestyle as they would like (Murray et al 2000; Forfar et al 2003). Exacerbations of asthma can be frightening and disruptive for child and family, and often necessitate hospital admission. Asthma affects QoL in children not just because of physical impairment, but also because of effects on emotional well-being (Annett 2001; Juniper et al 1999; van den Bermt et al 2010).

Chronic inflammation in the airways of children with asthma causes structural changes, known as airway remodelling, which can result in permanent damage. Although its long-term clinical implications are unknown, airway remodelling may precede life-long respiratory illness, as many children with asthma develop chronic respiratory problems in adulthood. Although there is no test for evaluating permanent structural lung damage in children, the degree of obstruction, at a particular point in time, can be measured using physiological markers, of which the two that are most commonly used in practice are Forced Expiratory Volume in one second (FEV1) and Peak Expiratory Flow Rate (PEFR). In order to measure FEV1, a child must exhale into a spirometry machine, via a mouthpiece, as fast as possible, for as long as they can. In order to measure PEFR, a child must exhale a short, sharp breath into a handheld device. Other pathophysiological effects of asthma can also be measured. Exhaled Nitric Oxide (ENO) is a marker of the degree of airway inflammation, and tests of bronchial hyper-responsiveness can give an indication of the sensitivity of a child’s airways. These tests are less frequently used in clinical practice, and tend to be measured in specialist centres.
Although mortality from asthma is rare in children, the International Study of Asthma and Allergies in Childhood (ISAAC), which is the largest epidemiological study of asthma to date, has found that the number of deaths attributable to childhood asthma, in several countries, increased between 1995 and 2002 (Anderson et al 2008). Worryingly, many children who die from asthma were previously considered to have mild or moderate illness (Robertson et al 1992).

1.6.2 Clinical trials of regular therapies for children with asthma

Until the 19th Century, asthma was regarded as a symptom of other respiratory ailments, rather than an illness in itself. Until the middle of the 20th Century, constriction of the airways was felt to be the underlying pathological feature of asthma, and treatments focussed on bronchodilation, with the aim of providing short-acting symptomatic relief.

The major advance in the understanding of the pathogenesis of asthma came in the mid 20th Century, when it became increasingly recognised that the major underlying factor was inflammation. As a result of this, asthma came to be considered a chronic condition, with inflammation occurring even when a patient may be free of symptoms. This led to the development of treatments that could modulate the underlying disease process. Therapies for asthma can, therefore, be classified as relievers, which provide symptomatic relief, or preventers, which aim to control symptoms.
and reduce the risk of exacerbations. This thesis relates to clinical trials of preventer therapies, rather than bronchodilators or other acute treatments for exacerbations.

By far the most widely used and successful preventative therapy, in both adults and children with asthma, has been inhaled corticosteroids (ICS). Steroid medication was first used for asthma in the 1950s, in an oral preparation, and was investigated, in a placebo-controlled RCT in children, which concluded that it effectively controlled symptoms and improved PEFR (Kennedy 1956). The need for inhaled preparations of steroids was initially driven by concerns that long-term use of oral cortisone was associated with a high incidence of serious systemic side effects. The development of inhaled steroids was made possible by the invention of the metered dose inhaler, in the late 1950s. Early trials of of Beclomethasone Diprropionate (BDP), the first ICS to be developed, evaluated whether it could control symptoms and improve PEFR, and also whether it could enable a reduction in the dose of oral steroid required (Godfrey, & König 1974). At around the same time, cromolyn medications, which modulated the inflammatory process, were evaluated in RCTs that measured similar outcomes (Smith, & Devey 1968).

As other types of ICS were developed, and methods for delivering them were refined, they replaced oral steroids as the main regular therapy for children with asthma (Chu, & Drazen 2005). The landmark RCT of ICS, in the 1980s, aimed to evaluate whether they could reduce bronchial hyper-responsiveness, which had been recognised as an important consequence of
airway inflammation (Haahtela et al 1991). Studies in the 1990s demonstrated the substantial burden of exacerbations of asthma, in terms of economic cost (Smith et al 1997), and damage caused to the airways (Barnes 1989). In the late 1990s, two landmark RCTs evaluating ICS in asthma, in children (Childhood asthma management program research group 2000), and children and adults (Pauwels et al 2003), demonstrated that they were effective at preventing exacerbations, as well as controlling symptoms and improving lung function.

Since their introduction in the 1970s, ICS have been widely prescribed, often at high doses, for children with asthma, but concerns remained that this formulation of steroids could also be associated with systemic side effects, albeit less severe than in those children who had been using oral cortisone. This was a major driver for the development of alternatives for ICS, and adjunct therapies that could be used as steroid-sparing agents.

Long-acting beta-agonists (LABA), which act primarily to reduce bronchoconstriction, and Leukotriene receptor antagonists (LTRA), which target specific aspects of the immune process, are the most widely used adjunct therapies in asthma. The main RCTs in children measured the effects of these drugs on lung function, symptoms, and short-term asthma control as primary outcomes (Bisgaard 2000; Russell et al 1995; Lazarus et al 2001; Lemanske Jr et al 2001).

There are problems with the existing evidence for therapies for children with asthma. In some situations, such as with LTRA, the majority of the evidence used to guide treatment in children has been extrapolated from
trials in adults. Furthermore, when systematic reviewers attempt to synthesise the results of these trials, they are faced with problems of non-uniform outcome measurement and reporting, which make this task difficult, if not impossible.

Largely because of these problems, even though ICS, LTRA and LABA are established therapies, that are known to do more good than harm, many questions remain unanswered about the best way to treat children with asthma. There is debate about who should administer asthma care, which treatments should be considered in the first instance, which should be second-line or additional agents, which devices should be used to deliver inhaled therapy, whether behavioural management of asthma is effective, and whether alternative homeopathic medicines are safe and helpful (James Lind Alliance 2007). There are also newer therapies such as Omalizumab, which modulate the immune response in asthma, and protect against the damage it causes, and these will require evaluation in large scale, late phase RCTs.

Two recent initiatives should increase the number and quality of such trials in children. Firstly, large paediatric research networks, such as the UK Medicines for Children Research Network (MCRN) have been set up across Europe. The problems with recruiting children into clinical trials, which were mentioned earlier, have necessitated the need for collaboration between centres, and initiatives such as the MCRN will facilitate the design and conduct of such studies across the UK. Secondly, there have been changes in the regulations on the licensing of medicines for children. These
were in response to the observation that many medications used commonly in children were unlicensed in children, because they had not been tested in the paediatric population (Conroy et al 2000; Bücheler et al 2002; Smyth, & Edwards 2006). Reasons for this lack of RCTs in children include ethical and practical complexities of conducting clinical research in a paediatric population (Smyth, & Weindling 1999), and the smaller financial gain that pharmaceutical companies receive from marketing medications in children, as compared with adults (Cohen et al 2007).

In the United States, in the late 1990s, the FDA Modernisation Act of 1997, the Best Pharmaceutical for Children Act, and the Paediatric Research Equity Act provided marketing incentives for pharmaceutical companies if they conducted FDA-approved RCTs in children (Hawcutt, & Smyth 2008). In 2007, The European Union (EU) Paediatric Regulation was passed, which offers extended periods of market patent for companies who conduct EMA-approved RCTs in children. Key to the European legislation is the Paediatric Investigation Plan (PIP), submitted by pharmaceutical companies to the EMA early in the drug development and evaluation process, that details the research and development that they will conduct to evaluate the efficacy and safety of a drug in children.
1.6.3 The need for research to identify the most appropriate outcomes for clinical trials in children with asthma

In order to answer important research questions, researchers and clinicians must take advantage of the initiatives described above, and conduct large-scale multi-centre RCTs in children with asthma. One of the factors that will determine the success of these trials is whether or not they measure the right outcomes. It can be difficult to know which outcomes are most important for evaluating treatments in childhood asthma. Some of the reasons for these problems were highlighted in Section 1.5.

The effects of asthma are diverse, and without research to identify which of these are important in clinical practice, it is not clear which should be measured as outcomes in clinical trials. Outcomes can measure the effects that interventions have on asthma at the level of the airways, such as measures of lung function, bronchial hyper-responsiveness or inflammation. Another approach is to assess the effects of these biological effects on the clinical features of asthma, such as symptoms or exacerbations. Yet another approach could be to measure the effects that the clinical features have on a child’s life, or their family, such as outcomes that reflect functional status or QoL. Further difficulties arise due to the complex natural history of asthma, for example it is unclear whether it is sufficient to measure short-term outcomes, or whether it may be more appropriate to measure lasting, long-term outcomes.
Generating a comprehensive picture of whether a child’s asthma has improved after starting an intervention can also be difficult because of methodological issues relating to data collection. Lung function tests are difficult to measure in young children, and diaries may be too complicated for children to complete. There are a variety of tools for measuring outcomes, especially QoL, and some of these may not have been validated (van den Bemt et al 2010).

When considering which outcomes to measure, trialists must also consider that physiological outcomes do not correlate well with more patient-centred outcomes, such as symptoms and QoL, and therefore may not give a comprehensive view of whether an intervention is helpful from the patient’s perspective (Brand et al 1999). This may confuse the issue of which intervention may be best for a patient, because the relative efficacy of different asthma medications may depend on the outcomes that are examined. For example, in one crossover RCT in adults with asthma, fluticasone (an ICS) was superior to eformoterol (a LABA) for clinic lung function and bronchial hyper-responsiveness, eformoterol was more effective for reducing the need for night-time reliever inhalers, and the drugs were equivalent for symptoms, PEFR, need for daytime reliever, QoL and patient assessments of asthma control (Jenkins et al 2005). Without knowledge of which of these outcomes is the most important, it can be difficult to know which medication is most appropriate for the patient. This was highlighted as a particular cause for concern by clinical and research experts at an international meeting in 2008, in which the
priorities for the improvement of asthma management and research were discussed (Holgate et al 2008).

Given how crucial the selection of appropriate outcomes is for the success of clinical trials, research to identify which outcomes are most important can be useful for a variety of people. Trialists from academia and industry who design and conduct RCTs, clinicians, patients and parents, and systematic reviewers, who synthesise the results of similar trials, will all benefit from a core outcome set that includes important outcomes. The purpose of this thesis is to examine the methodology of identifying the most appropriate outcomes for clinical trials, and to determine which outcomes are important in childhood asthma.

1.7 Structure of this thesis

In Chapter 2, a systematic review of studies that determine which outcomes to measure in clinical trials in children is described. The purpose of this review was to identify specialties in which such work had been conducted, and to examine the methodology used.

In Chapter 3, a systematic review of RCTs in childhood asthma, published between 1988 and 2008, is described. This work was conducted because, in the review in Chapter 2, we found no robust core outcome set for childhood asthma. We, therefore, wished to identify which outcomes were measured in RCTs in asthma, investigate whether any outcome domains
were under-represented, and determine whether there was non-uniformity between studies. The main recommendation from this review was that a core outcome set should be developed for RCTs in childhood asthma.

In Chapter 4, a systematic review of studies that have used the Delphi technique, to design core outcome sets for clinical trials, is described. This review highlighted important considerations when designing and reporting such studies.

In Chapter 5, a study to determine which outcomes are most important in childhood asthma is reported. The methodology for this study, which reflects the recommendations from the systematic review described in the previous chapter, comprised a Delphi study involving paediatricians and respiratory nurses in the UK, and parents of young people with asthma, to determine the most important outcomes to these groups.

In Chapter 6, an outline of the possible future directions of our research is presented.
Chapter 2

A systematic review of studies that aim to determine which outcomes to measure in clinical trials in children

2.1 Background

When designing a clinical trial, the choice of which outcomes to measure is crucial. The selection of inappropriate outcomes can lead to wasted resources or misleading information that overestimates, underestimates, or completely misses the potential benefits of an intervention. Examples of these problems are well documented (Fleming, & DeMets 1996; Holloway, & Dick 2002).

It can be difficult to know which outcomes should be measured in clinical trials, because illnesses can have several effects, which could each improve after starting an intervention, and could therefore be selected as outcomes. There may be a lack of methodological research to determine which of these are most appropriate for clinical trials in a given condition.

In light of these complexities, some work has been conducted, notably in the field of rheumatology, to determine the outcomes that researchers and clinicians feel should be measured in clinical trials (Tugwell et al 2007). A few studies, conducted in adults, have also aimed to identify the most
important outcomes from the perspective of patients and families (Kirwan et al 2005; Mancuso et al 2002; Arnold et al 2008). Research such as this should improve the likelihood that important outcomes are measured in clinical trials. Standardising outcomes across trials also makes it easier to synthesise the results of otherwise similar studies, by ensuring that outcomes are measured and reported in a uniform manner, and reduces the risk of selective reporting.

This chapter describes a systematic review of studies that address the question of which outcomes to measure in clinical trials in children, under 16 years of age. We have restricted this review to studies identifying outcomes for clinical trials in children because there is increasing recognition that children are not merely “small adults,” and the methodology of conducting research in this age group should be tailored accordingly. The outcomes that are appropriate for clinical trials in children may also, justifiably, differ from those relevant to clinical trials in adults. We anticipated that one way of determining which outcomes to use, in addition to the consensus techniques described above, may be to ascertain the opinions of both children and their parents regarding outcomes that they think are important. This process poses unique challenges that may not be relevant when conducting similar research in adults with a particular condition, so we deemed it appropriate to specifically review studies pertaining to the selection of outcomes, or outcome domains, in clinical trials in children.
2.2 Aims

1. To identify studies which have been conducted to determine which outcomes or domains to measured in clinical trials in children.

2. To identify the methodological techniques utilised in these studies.

3. To assess the quality with which the studies were conducted and reported.

2.3 Methods of the systematic review

2.3.1 Inclusion and exclusion criteria

Studies that developed or applied methodology for determining which outcome domains or outcomes should be measured in clinical trials in children younger than 16 years of age were eligible for inclusion, as were systematic reviews of these studies.

The following studies were excluded from the review:

1. Studies that did not specifically state that the outcomes were considered for use in a paediatric population.
2. Studies relating to instruments, tools, scales, scores, and definitions, which address how outcomes should be measured, rather than which outcomes to measure.

3. Studies relating to clinical trials that assess interventions given to adults, by measuring outcomes in children, for example the selection of neonatal outcomes to assess care given to their mothers.

2.3.2 Identification of relevant studies

In order to optimise the chances of identifying all relevant studies, a variety of medical literature databases were searched. These are described below:

The Cochrane Library

http://www3.interscience.wiley.com/cgi-bin/mrwhome/106568753/HOME

This collection of databases contains primary and secondary biomedical and methodological research. The search was conducted simultaneously in the Cochrane Database of Systematic Reviews, Cochrane Database of Abstracts of Reviews of Effects, Cochrane Database of Methodology Reviews, Cochrane Methodology Register and Cochrane Central Register of Controlled Trials.
**Medline**


This database contains citations and abstracts from over 5000 journals relating to medicine, biomedical sciences and allied health professions. Coverage is from 1966 onwards.

**Cumulative Index of Nursing and Allied Health Literature (CINAHL)**

http://www.cinahl.com/library/library.htm

This database of around 3000 journals, books, dissertations, conference proceedings and websites, relates to research in nursing, biomedical sciences and the allied health professions. Coverage is from 1981 onwards.

**SCOPUS**

(http://www.scopus.com/)

This platform is used to search a variety of databases. EMBASE is one such database, which contains over 700 journals relating to pharmacology and biomedical sciences. Searching both EMBASE and Medline increases the chances of identifying all available literature (Wilkins et al 2005; Suarez-Almazor et al 2000).
2.3.3 Designing an appropriate search strategy

The design of the search strategy was an iterative process, in which search terms were modified to include synonyms and variations identified in retrieved publications (Higgins, & Green 2008). The process of designing the search strategies is described below (Jones, & Smyth 2004). The search strategies are shown in Appendix 1.

Synonyms were listed for the four specific terms “outcomes”, “methodology”, “clinical trials” and “children”, which reflected the aim of the review. Synonyms within in each of these four word lists were combined using the Boolean operator OR, so if a study citation included any of these terms in its title, abstract, or indexing words (either keywords specified by the authors, or MeSH words nominated by Medline librarians), it would be identified. To narrow the search, the four lists of synonyms were combined using the Boolean operator AND, so that only those studies which included at least one term from each list would be identified.

Truncation and wildcards were used to improve the sensitivity of the search. Truncation, denoted by an asterix, identifies different derivations of search terms. For example child* would identify child, children, childhood etc. Wildcard characters, denoted by a question mark, account for spelling variations. For example p?ediatric would identify the terms paediatric and pediatric.
2.3.4 Selecting studies for inclusion in the review

Each identified abstract was screened to assess eligibility. To reduce the chance of missing relevant studies, this process was performed twice. Full texts of potentially relevant articles were assessed with regard to the predefined eligibility criteria.

To check this process, a second reviewer independently screened a database that comprised all the abstracts for which full text had been obtained, and a selection of those that had been rejected. A sample, rather than the complete set, was selected due to resource constraints.

A list of identified studies was emailed to the Clinical Study Group (CSG) members of the MCRN, who were asked to suggest other potentially eligible studies, published or unpublished. The speciality-specific CSGs constitute a multidisciplinary group of experts with a strong interest in the planning of clinical trials.

2.3.5 Data Extraction

From each study, the following data were extracted by one reviewer (IS), and checked by the second reviewer:

1. The condition for which the outcome domains or outcomes were considered

2. A description of the method
3. People involved in selecting outcome domains or outcomes

4. The outcome domains and outcomes that were selected

5. The geographical setting of the study. This was ascertained either by reading the text or, where listed, the names and institutions of people involved in the collaborations

2.3.6 Assessing the quality of included studies

The methodological quality of the studies was assessed by IS, and checked by the second reviewer. Reporting quality was assessed in terms of whether the study methods were reproducible.

No validated assessment tool for critically appraising consensus statements existed at the time of this review, so two experts, one with experience of qualitative research, and the other with experience of consensus statements, were asked to advise on this assessment. To assess the quality of systematic reviews, the Critical Appraisal Skills Programme Systematic Review Appraisal tool was used (http://www.phru.nhs.uk/Pages/PHD/resources.htm).

2.3.7 Data Analysis

Studies were described narratively, and results presented in textual format.
2.4 Results

2.4.1 Study selection

The process of identifying studies is summarised in Figure 1 (shown on the next page).
Figure 1: Flowchart of the review: identification of studies that aim to determine which outcomes to measure in clinical trials in children.
The initial database search identified 8,889 potentially relevant abstracts, of which 8,819 were excluded. Of 70 articles that were retrieved in full, 25 were included in the full review. These described studies conducted by 13 separate collaborative groups. In addition, the members of the MCRN CSGs suggested 13 other articles. One of these summarised the work of a collaboration that was identified by the literature search, but did not itself describe the methodology used in sufficient detail to warrant inclusion in the full review, so is added as an additional reference (Giacoia et al 2006). The other 12 suggested studies were not deemed relevant.

In total, therefore, 57 full-text articles were reviewed and subsequently excluded. Of these studies, 19 were excluded because the authors did not use methodology for selecting outcomes (e.g. a review article based on personal opinion), 18 related to how to measure outcomes, rather than which outcomes to measure, 10 made no mention of outcome selection, 6 did not specifically state that the outcomes which were selected were relevant to children, and 4 described consensus statements relating to clinical practice rather than clinical trial design. The reasons for exclusion of each individual study are presented in Appendix 2.

2.4.2 Agreement between reviewers

The second reviewer was provided with a database of 100 abstracts. These included a randomly selected sample of 30 abstracts that had been excluded at the initial screening stage, and 70 for which full text had been retrieved.
The second reviewer agreed with the exclusion of all 30 abstracts that were rejected at the initial screening stage. Of the 70 abstracts for which full text had been obtained, the second reviewer agreed with 61, and disagreed with 9. After discussion, it was agreed that all nine should be retrieved in full based on the abstract. Of these, eight were subsequently excluded, after reading the full text, and one was included. Following full text review there was complete agreement between reviewers about the 25 included and 45 excluded abstracts. Subsequently, there was also complete agreement regarding the data extraction and methodological assessment of the studies.

2.4.3 Description of included studies

The 25 articles included in the full review represented the work of 13 collaborative groups, which are listed in Table 1. A detailed summary of each study is included in Appendix 3.
Table 1 List of studies included in the review

<table>
<thead>
<tr>
<th>Speciality</th>
<th>Condition</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Critical care</td>
<td>Sepsis</td>
<td>Goldstein 2005</td>
</tr>
<tr>
<td>Dentistry</td>
<td>Dental restoration</td>
<td>DeRouen 2002</td>
</tr>
<tr>
<td>Gastroenterology</td>
<td>Crohn’s disease</td>
<td>Griffths 2005</td>
</tr>
<tr>
<td>Transplant medicine</td>
<td>GVHD</td>
<td>Pavletic 2006</td>
</tr>
<tr>
<td>Neonatology c</td>
<td>Apnoea</td>
<td>Finer 2006</td>
</tr>
<tr>
<td></td>
<td>Cardiac instability</td>
<td>Short 2006</td>
</tr>
<tr>
<td></td>
<td>Postoperative cardiac dysfunction</td>
<td>Roth 2006</td>
</tr>
<tr>
<td></td>
<td>Seizures</td>
<td>Clancy 2006</td>
</tr>
<tr>
<td>Neurology</td>
<td>Infantile spasms</td>
<td>Lux 2004, Osborne 2001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(on behalf of West Delphi Group)</td>
</tr>
<tr>
<td>Psychiatry</td>
<td>Bipolar affective disorder</td>
<td>Carlson 2003</td>
</tr>
<tr>
<td></td>
<td>Nonepileptic seizures</td>
<td>La France 2006</td>
</tr>
<tr>
<td>Respiratory medicine</td>
<td>Asthma</td>
<td>Smith 1996</td>
</tr>
<tr>
<td></td>
<td>Cystic fibrosis</td>
<td>Ramsey 1994</td>
</tr>
<tr>
<td></td>
<td>Systemic Lupus</td>
<td>Ruperto 2003, Ruperto 2004, Ruperto 2006 (on behalf of PRINTO e)</td>
</tr>
<tr>
<td></td>
<td>Erythematosis/</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dermatomyositis</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rheumatoid arthritis</td>
<td>Giannini 1997</td>
</tr>
</tbody>
</table>

Footnotes and legend to Table 1:
a Concerned the selection of outcomes for use in a specific clinical trial
b Concerned the selection of outcomes which could be used in clinical trials of children with clinical features of an underlying condition
c Collaboration working on behalf of the Neonatal Drug Development Initiative (NDDI)
d IMACS = International Myositis Assessment and Clinical Studies Group
e PRINTO = Paediatric Rheumatology International Trials Organisation
Four groups (Carlson et al 2003; Goldstein et al 2005; LaFrance et al 2006; Lux, & Osborne 2004) and the five eligible studies within the Neonatal Drug Development Initiative (NDDI) (Anand et al 2006; Clancy 2006; Finer et al 2006; Roth et al 2006; Short et al 2006) considered outcomes while addressing wider clinical trial design issues.

Six groups (Griffiths et al 2005; Ramsey, & Boat 1994; Pavletic et al 2006; Giannini et al 1997; Miller et al 2001; Ruperto et al 2003) aimed to reach consensus specifically around outcome selection and measurement.

One group (Smith et al 1996) aimed to ascertain the opinions of clinicians, about outcomes which could be measured in clinical trials in children with asthma, but did not aim to reach a consensus around this topic.

One group (DeRouen et al 2002) addressed the question of which outcomes to measure in a specific trial in paediatric dental restoration. No systematic reviews of these, or other relevant studies, were identified.

Eight groups were based in the USA (Carlson et al 2003; DeRouen et al 2002; Goldstein et al 2005; Griffiths et al 2005; LaFrance et al 2006; Pavletic et al 2006; Ramsey, & Boat 1994; Giacoia et al 2006). One group was based in Europe, but included participants from several countries in Europe, North America, and Asia (Lux, & Osborne 2004). One group was based in Australasia (Smith et al 1996). The three rheumatology collaborations (Giannini et al 1997; Miller et al 2001; Ruperto et al 2003) involved clinical researchers in the US and Europe.
2.4.4 Participants involved in the studies

Clinicians

All 13 groups included clinicians with expertise in treating the conditions for which outcomes were being selected.

Research experts

All groups included participants with experience of conducting clinical research in the condition for which outcomes were being selected. In addition to clinical research experts, some groups also included biostatisticians and epidemiologists. It is difficult to quantify exactly how many groups employed such personnel, because they may have been referred to as “research experts”.

Patients or parents

Three groups involved parents of children with medical conditions, but none involved children directly. One study (Carlson et al 2003) involved “representatives of families with affected children” in discussions about outcomes for clinical trials of children with bipolar affective disorder. One study report (Pavletic et al 2006) acknowledges “patients and patient and research advocacy groups”. The level of involvement of these patient representatives was not described in either article. One study report (Miller
et al 2001) acknowledges funding and administrative support given by two named patient support group leaders.

**Representatives from the pharmaceutical industry and drug regulatory authorities**

Three groups included participants whose interest in clinical trials related to drug licensing and marketing. The Neonatal Drug Development Initiative (NDDI) is described as a collaboration between the FDA and “neonatal experts and colleagues, representing industry and academia” (Giacoia et al 2006). Carlson (2003) mentions the involvement, in a consensus statement about outcomes for clinical trials of children with bipolar disorder, of “pharmaceutical industry sponsors with an interest in mood stabilizer products, staff of the FDA and their counterparts from regulatory agencies in Canada and the European Union”. Participants in a consensus conference to identify outcomes for clinical trials in Cystic Fibrosis included “representatives from both the Cystic Fibrosis Foundation and the U.S. Food and Drug Administration” (Ramsey, & Boat 1994).
2.4.5 Methods used to select outcomes

The following techniques were used to determine which outcomes should be measured in clinical trials in children with specific conditions.

**Delphi technique**

The Delphi technique (Dalkey 1969) is a method of reaching a consensus opinion, in which sequential questionnaires are answered, anonymously, by participants. After each questionnaire, the group response is fed back to participants. Three groups utilised this method as follows.

The West Delphi group (Lux, & Osborne 2004) used the Delphi technique to identify which outcomes should be measured in clinical trials of children suffering from infantile spasms. The process was conducted, by email, over six rounds. In the first round 133 invited participants, of whom 42 responded, were asked multiple-choice questions covering various aspects of clinical trial design, including outcomes. In the second round, multiple-choice questions were asked about these topics. At this stage participants were also invited to comment and provide their personal opinions regarding outcomes. In the third round, statements were formulated from earlier responses. Participants were invited to answer whether they agreed or disagreed with these statements. In the fourth round, these statements were modified and participants commented on their suitability and content. The final two rounds involved the formulation of the final consensus statement and study report.
The PRINTO group (Ruperto et al 2003) used the Delphi technique, over two sequential questionnaire-based surveys, to identify which outcomes should be measured in clinical trials of children with SLE. In the first questionnaire they asked 267 participants to indicate up to 10 variables they judged as clinically important. In the second questionnaire, the facilitators listed those indicators that had been suggested by at least ten responders, and the participants ranked, in order, their top ten choices.

The IMACS group (Miller et al 2001) used a Delphi technique to develop a core set of outcome domains and outcomes for use in clinical trials in children with inflammatory myopathy. The actual process itself is not described in detail in the article, but authors stated that the group consisted of “more than 100” members.

Nominal group technique

Nominal Group Technique (NGT) involves structured face-to-face discussion, followed by a vote on the options presented. Two groups utilised this technique.

The PRINTO group (Ruperto et al 2003) used NGT to select outcomes from those identified by the Delphi technique described earlier. The NGT exercise had five objectives, which were tackled by a group of 40 participants: (1) to classify the proposed outcomes into “domains”; (2) to classify the outcomes into “concepts of disease activity”; (3) to select the outcome domains that should be measured in clinical trials; (4) to select the
outcomes that should be used to measure these domains; and (5) to discuss specific design issues of the prospective validation phase of the study.

Giannini et al (1997) used NGT to select, from a set of potential outcomes, a preliminary core set of six. The process used is not described in further detail in the study. The initial list of potential outcomes had been identified by sending a questionnaire to a 16-member advisory council.

Semistructured discussion

Eight groups, including the studies within the NDDI collaboration, came to consensus by discussion at meetings or workshops (Goldstein et al 2005; DeRouen et al 2002; Griffiths et al 2005; Pavletic et al 2006; Carlson et al 2003; LaFrance et al 2006; Ramsey, & Boat 1994; Anand et al 2006; Clancy 2006; Finer et al 2006; Roth et al 2006; Short et al 2006).

Questionnaires

Smith et al (1996) sent questionnaires to 39 health care professionals and researchers with expertise in asthma to ask which outcomes they would use for a variety of clinical, research, and public health scenarios, including questions about which outcome they would use in clinical trials of acute and preventative asthma medication. Three other groups (Giannini et al 1997; Lux, & Osborne 2004; Ruperto et al 2003) used questionnaires as
part of the process of ascertaining the opinions of experts, mainly in the preliminary phases of the consensus process.

2.4.6 Outcomes and domains that were selected

Authors of the included studies either categorised outcomes into disease-specific domains, or presented a list of un categorised outcomes. Having listed the outcomes selected in each study, we categorised them into the following six domains, which were felt to be applicable to all paediatric conditions: disease activity; complications of the disease; adverse effects of therapy; functional status; social outcomes, family outcomes and QoL; health resource utilisation. The outcomes selected by each group are shown in Appendix 4, categorised by these domains.

2.4.7 Assessment of the quality of studies

Reporting of methodology

Of the four collaborations who used structured consensus formulation techniques, namely one or both of NGT or Delphi Technique, three described the process of reaching consensus clearly in the study report (Giannini et al 1997; Lux, & Osborne 2004; Ruperto et al 2003). Each of the eight collaborations who came to a consensus after structured discussion, but without using Delphi technique or NGT, described the discussions in some detail, but the actual process of how consensus was
reached was generally not reported. The methodology used in the study that used a single questionnaire based survey (Smith et al 1996) was described in sufficient detail to be able to repeat the study. All groups described the background of their participants. Only two, however, described in detail the process by which it was decided which particular individuals would be involved (Lux, & Osborne 2004; Smith et al 1996).

The reporting quality of the methodology used in the individual studies is summarised in Table 2.
<table>
<thead>
<tr>
<th>Study</th>
<th>Process of reaching consensus described in detail?</th>
<th>Description of the participants</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Expertise of participants described?</td>
<td>Selection of individuals described?</td>
</tr>
<tr>
<td>Anand 2005, 2006</td>
<td>Discussion points are described. Actual process of reaching consensus is not described.</td>
<td>Yes</td>
</tr>
<tr>
<td>Carlson 2003</td>
<td>Discussion points are described. Actual process of reaching consensus is not described.</td>
<td>Yes</td>
</tr>
<tr>
<td>Clancy 2006</td>
<td>Discussion points are described. Actual process of reaching consensus is not described.</td>
<td>Yes</td>
</tr>
<tr>
<td>De Rouen 2002</td>
<td>Discussion points are described. Actual process of reaching consensus is not described.</td>
<td>Yes</td>
</tr>
<tr>
<td>Finer 2006</td>
<td>Discussion points are described. Actual process of reaching consensus is not described.</td>
<td>Yes</td>
</tr>
<tr>
<td>Giannini 1997</td>
<td>Process of coming to consensus is clearly described</td>
<td>Yes</td>
</tr>
<tr>
<td>Goldstein 2005</td>
<td>Discussion points are described. Actual process of reaching consensus is not described.</td>
<td>Yes</td>
</tr>
<tr>
<td>Griffiths 2005</td>
<td>Discussion points are described. Actual process of reaching consensus is not described.</td>
<td>Yes</td>
</tr>
<tr>
<td>La France 2006</td>
<td>Discussion points are described. Actual process of reaching consensus is not described.</td>
<td>Yes</td>
</tr>
<tr>
<td>Lux 2004, Osborne 2001</td>
<td>Process of coming to consensus is clearly described</td>
<td>Yes</td>
</tr>
<tr>
<td>Miller 2001, Rider 2002, 2003, 2004; Oddis 2005</td>
<td>Techniques used for coming to consensus are described. The actual process of coming to consensus is not described in detail.</td>
<td>Yes</td>
</tr>
<tr>
<td>Pavletic 2006</td>
<td>Discussion points are described. Actual process of reaching consensus is not described</td>
<td>Yes</td>
</tr>
<tr>
<td>Ramsey 1994</td>
<td>Discussion points are described. Actual process of reaching consensus is not described</td>
<td>Yes</td>
</tr>
<tr>
<td>Roth 2006</td>
<td>Discussion points are described. Actual process of reaching consensus is not described.</td>
<td>Yes</td>
</tr>
<tr>
<td>Ruperto 2003, 2005, 2006</td>
<td>Process of coming to consensus is clearly described</td>
<td>Yes</td>
</tr>
<tr>
<td>Short 2006</td>
<td>Discussion points are described. Actual process of reaching consensus is not described.</td>
<td>Yes</td>
</tr>
<tr>
<td>Smith 1996</td>
<td>Process of ascertaining opinions from experts is described</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Techniques Used to Validate Outcomes

Two groups made some attempt to prospectively validate the outcomes they had selected. In one study (Giannini et al 1997), the multicollinearity and redundancy of a core set of outcomes for use in clinical trials of children with rheumatoid arthritis was assessed in a group of children in a clinical practice, and using a database from a previous observational cohort study. The acceptability of the core set of outcomes to a wider group of clinicians was assessed by sending a questionnaire to an international selection of rheumatologists, seeking their reactions to the outcomes.

The PRINTO group (Ruperto et al 2003) prospectively validated the core set of outcomes they had produced for clinical trials of children with SLE. This was done by measuring the outcomes in patients in a clinical outpatients setting who were being started on new modalities of treatment for their condition. In this way the authors aimed to replicate a clinical trial setting, and assess the feasibility, discriminative ability, validity, and internal consistency of the core set of outcomes.

Both of these groups, and the IMACS group (Miller et al 2001) also developed “definitions of improvement,” based on the degree of change within each outcome, which could be used as a dichotomous index in clinical trials to determine whether patients had benefited from the treatment they had received. This was done in all three studies by developing a set of “paper patient profiles,” and asking clinicians whether or not they thought the patient had improved. A set of potential definitions
of improvement was narrowed down to a final definition by using various consensus techniques.

2.4.8 Additional studies identified after this systematic review was conducted

We identified two studies that met the eligibility criteria for inclusion in this systematic review, but were published after its completion.

PedIMMPACT

The IMMPACT (Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials) collaboration has developed core outcome sets for use in clinical trials that assess treatments for patients with pain. One study conducted by this group relates to outcomes in clinical trials of children suffering from pain (McGrath et al 2008).

This study involved 26 participants, of whom 23 were based in North America, and 3 in Europe. These included representatives from the clinical research community, drug regulatory authorities, and the pharmaceutical industry. The group conducted a two-stage Delphi process in which important outcome domains, and possible outcomes within these domains, were suggested and scored by participants. The results of this Delphi process were discussed at a consensus meeting at which, after semi-
structured discussion, the group reached consensus about a core set of outcomes.

The core outcomes suggested for children with acute pain were pain intensity, global judgment of satisfaction with treatment, symptoms and adverse events, physical recovery, emotional response, and economic factors. The core outcomes suggested for children with chronic pain were pain intensity, global judgment of satisfaction with treatment, symptoms and adverse events, physical functioning, emotional functioning, role functioning, sleep, and economic outcomes.

**Asthma**

In July 2009, a report was published which described a consensus conference relating to outcomes in clinical trials of patients, aged 6 years and older, with asthma (Reddel et al 2009). The aims of the consensus group were to define the outcomes that are most important for measuring asthma control, to determine the most appropriate way to measure these, and to define the term ‘exacerbation’.

Seven working groups each discussed one of the following topics: exacerbations; diary data; physiological measures; composite scores; biomarkers; indirect measures; or quality-of-life questionnaires. The discussions were informed by literature reviews around each of the topics. After each working group had formulated some guidelines, these were circulated to the other groups for discussion and comments. One other
group, comprising two paediatricians with experience of clinical research, offered advice on each working group’s suggestions, from the perspective of clinical trials in children.

Composition of the group “was intended to represent a broad spectrum of clinical expertise and clinical trial experience”, but the areas of expertise of the participants is not described in further detail. In total, 25 clinicians were involved, including the paediatricians. The FDA and EMA provided, between them, three representatives. Pharmaceutical companies were not involved in the discussions, but were invited to suggest “written submissions”, of which three were circulated to the group members. It is unclear what they suggested. No patients or families were involved in the study. Participants were from the USA, Europe, Australasia and South Africa.

The core outcomes were divided into those relating to current clinical control, and those relating to future risk. Outcomes were classified as essential, desirable and optional.

The ‘essential’ core outcomes relating to current control were: Symptom-free days; Reliever use; Composite scores (eg Asthma Control Questionnaire); Exacerbations; Quality of life

The ‘desirable’ outcomes relating to current control were: FEV1; daily symptoms, reliever use and lung function, as recorded in a patient diary; corticosteroid use and health care utilisation.
The ‘optional’ outcomes relating to current control were biomarkers, airway hyperresponsiveness, and post-bronchodilator FEV1.

The ‘essential’ core outcomes relating to future risk were: Composite scores (e.g., Asthma Control Questionnaire); Exacerbations; Post bronchodilator FEV1 (as a measure of lung decline); treatment side effects, and Pre-bronchodilator FEV1 (as a predictor of exacerbations).

The ‘desirable’ outcomes relating to future risk were: daily symptoms, reliever use and lung function, as recorded in a patient diary; health care utilization (e.g., corticosteroid use, ER visits, hospitalizations); mortality due to asthma; Airway hyperresponsiveness (as a predictor of future risk); Biomarkers (as predictor of future risk).

No ‘optional’ outcomes relating to future risk were suggested.

2.5 Discussion

We systematically reviewed the work of 13 groups who addressed the issue of selecting outcomes for use in paediatric clinical trials. In many paediatric specialties, such work has not been conducted. The quality of the studies, with regard to the methodology used, and the level of detail with which the methods were reported, was variable.

We identified three methods used for reaching consensus, namely NGT, Delphi technique, and semi-structured discussion. Many groups used a
multidisciplinary approach to the problem of outcome selection, including researchers with experience of clinical trial design, statisticians, and clinicians. Some groups also involved representatives from industry or drug regulatory authorities, but the nature of their involvement is not evident from reading the reports.

As the aim of clinical trials should be to determine whether patients experience important benefits from an intervention, it was notable that we did not identify any studies that had directly asked children what they considered to be the most relevant outcome domains or outcomes. In the United Kingdom steps are being taken to involve consumers in medical research. A major initiative is the James Lind Alliance (JLA) (http://www.lindalliance.org/), a collaboration that aims to ascertain from a variety of people involved in clinical research, and patients, what they think are the most pressing research priorities for various conditions. Determining appropriate outcomes for paediatric studies is another area in which children and families can be involved in clinical trial design.

Robustness of the review

Our review was conducted in a rigorous, systematic manner. Two reviewers adhered to strict eligibility criteria to determine which studies should be included. Although the sample of excluded papers checked by the second reviewer represented a small proportion of all the ineligible studies, we concluded that agreement between the reviewers was adequate.
We determined that a smaller proportion of excluded studies would be sufficient for quality assurance, as compared with other systematic reviews in which the results of clinical trials are synthesised.

Although the 8,889 abstracts identified were screened twice, it may be possible that some relevant studies were missed. The types of studies that may not have been identified at this stage include clinical trials that did not describe in the abstract how the authors selected their outcomes, but subsequently in the full text may have mentioned the process used. It is also possible that some studies may have been missed by not searching the “grey” literature such as unpublished conference proceedings.

There were recurring features of the methodology and reporting quality of the consensus statements that may have compromised the scientific validity of the studies we identified. Most studies that described formation of a consensus statement did not explain in sufficient detail two key aspects of the process, namely the method used to select participants, and the process by which consensus was reached. Insufficient information was given to determine the level of involvement of certain groups involved in the research, particularly industry representatives, drug regulatory authority representatives, and parents of affected children.
**Implications for clinical trials**

If implemented, the studies we identified could improve the standardised measurement, of important outcomes, in clinical trials. Uniform selection of outcomes would make interpretation of results and comparison across trials simpler, hence making meta-analyses easier and more powerful (Clarke 2007). Disease-specific, universally agreed core sets of outcomes, that should be measured and reported in all clinical trials of a specific condition, regardless of statistical significance, have also been advocated as a solution to the common problem of ORB (Williamson et al 2005; Giannini et al 1997; Kirkham et al 2010).

**Implications for drug regulatory authorities**

As discussed in Chapter 1, there is a need for high quality paediatric clinical trials, and the development of Paediatric Investigation Plans (PIPs) is one of the changes in drug regulation in Europe that should facilitate this goal. The PIP, which is submitted to the EMA, is a detailed outline of research that would be needed to investigate the potential benefits and harms of medications for use in children. If a drug company were to be involved in the writing and implementation of a PIP, they would be eligible for marketing rewards in the form of prolonged patent protection and market exclusivity. When a PIP is submitted, the endpoints selected for the trial must be clearly stated and their appropriateness described (http://www.emea.europa.eu/htms/human/paediatrics/pips.htm).
The studies we have identified should be of use to people designing a PIP, and it is possible that the types of studies we have identified may become more common as drug companies seek to take advantage of the benefits of conducting high-quality clinical trials in children. The new standards for conducting clinical trials of investigational medicinal products set by the EMA aim to improve the quality of paediatric research. In order to obtain a license for a drug, it must be investigated according to these guidelines (http://www.emea.europa.eu/htms/human/humanguidelines/efficacy.htm).

In July 2007, of the 13 paediatric conditions identified in this review, the EMA Web site included guidelines for one (juvenile idiopathic arthritis) and a concept paper discussing the need for guidelines for another (cystic fibrosis).

2.6 Summary

We have reviewed 13 studies that address the process of selecting outcomes for clinical trials in children. This work should make it easier to determine which outcomes to measure in clinical trials. Although it is commendable that there are existing collaborations in several clinical areas, future work in this area may be improved by involving children and parents. The studies identified by this review will go some way to improving the quality of paediatric research, but further research is justified and needed.
One of the clinical areas in which we found no core outcome sets, designed using structured consensus technique, is childhood asthma. One possible implication of this may be that trials conducted in children with asthma may select and measure outcomes in a non-uniform manner. Another possible implication may be that, even if trialists are measuring the same outcomes, certain important effects of treatments are being overlooked and under-represented. The purpose of the review described in Chapter 3 is to identify whether these problems are evident in clinical trials of regular therapies for children with asthma.
Chapter 3

A systematic review to identify which outcomes are measured in randomised controlled trials in children with asthma

3.1 Background

Asthma in children is a major global health problem (Bateman et al 2008). It is an important cause of morbidity, mortality, and economic cost (Masoli et al 2004), it is the commonest chronic condition in industrialised countries (Beasley 1998), its prevalence is increasing (O’Connell 2004), and in many children it is a progressive condition that continues into adulthood (Martinez 2002).

Several outcomes are relevant when evaluating the efficacy and safety of treatments in asthma, and it may be difficult for researchers to know which of these should be measured in clinical trials. One of the studies included in the systematic review in Chapter 2 related to outcomes for RCTs in children with asthma (Smith et al 1996). As this study consulted only a limited group of paediatric experts, did not ask about all relevant domains, did not use structured consensus techniques, and did not involve children with asthma, or their parents, it does not provide a robust basis for recommendations about which domains and outcomes are most appropriate for RCTs of children with asthma. One other study, published after the
systematic review was completed (Reddel et al 2009) used a more structured approach to reaching consensus about which outcomes to measure in clinical trials, but is too recent to have been implemented yet. The absence of a core outcome set, designed using structured consensus techniques, may have led to non-uniformity between studies, and may also mean that certain outcome domains have been overlooked. The systematic review described in this chapter evaluates which outcomes and domains are measured in clinical trials in asthma, to assess whether such problems are prevalent.

### 3.2 Aims

1) To assess which outcomes had been measured in clinical trials of ICS, as regular therapy for children with asthma, between 1988 and 2007, in order to determine whether all domains were represented, and whether there was consistent selection of outcomes within these domains.

2) To determine whether the selection of outcome domains has changed over this period.

3) To determine whether domains measured in RCTs that exclusively involve children differ from those in studies involving both children and adults.

4) To determine whether domains measured in publically funded trials differed from those in trials funded by the pharmaceutical industry.
3.3 Methods of the review

3.3.1 Inclusion and exclusion criteria

In order to ensure that a group of similar studies was examined, this review included only RCTs with parallel group design that assessed ICS as a therapy to prevent symptoms or long-term effects of asthma in children. RCTs that only involved children younger than 18 years of age, and also studies that included children and adults, were eligible.

Crossover trials were excluded, because they are generally of acute interventions, the length of treatment in these studies is typically shorter, and the outcomes they measure may differ from those measured in parallel trials (Pocock 1983). In order to only include RCTs assessing long-term preventative therapy for asthma, studies with a treatment phase of less than one month were excluded. The review was restricted to studies published between January 1988 and December 2007.

3.3.2 Identification of studies

We searched the Cochrane Central Register of Controlled Trials (CENTRAL). This database comprises citations to reports of clinical trials, from Medline, EMBASE, and other sources, such as registries and journals that are not indexed in Medline. Because of the wide range of sources that
are covered by CENTRAL, it is regarded as the most comprehensive database of RCTs (Higgins, & Green 2008).

3.3.3 Search strategy

The process of designing and refining a search strategy is described in section 2.3.3. As the eligible studies were randomised controlled trials that assessed ICS as a therapy for children with asthma, the word lists that were generated included synonyms for the terms “children”, “inhaled corticosteroids” and “asthma”. The terms within each of these lists were combined using the Boolean operator OR, and then the word lists were combined using the operator AND. The full search strategy is included in Appendix 5.

3.3.4 Data extraction and quality assessment

From each report, the following data were extracted by one reviewer (IS):

1. All outcomes reported

2. If stated, the designated primary outcome, and whether this was described in sufficient detail, including the methods used to measure it, by whom, and when it was measured and analysed.
3. Other study features: year of publication; interventions compared; ages of children included; length of study treatment; single- or multi-centre; source of funding. The source of funding was identified either directly from a statement in the study report or, if this was not present, by contacting the authors using email.

4. Whether interventions were adequately masked. This was examined because the extent to which interventions were masked may affect the choice of outcomes, and whether they were measured objectively or subjectively. The adequacy of masking was categorised as follows.

Adequately masked:

Authors either clearly describe or imply, in the methods, how the allocated treatment was masked to the patient and family, medical caregiver, and trial personnel involved in measuring outcomes.

Inadequately masked:

Authors specifically state that the identity of the allocated treatment arm was not masked to at least one of the following: patient and family, medical caregiver, or relevant trial personnel involved in measuring outcomes.

Unclear:

It was unclear from study methods whether masking was adequate or not.
3.3.5 Data analysis and presentation

For each study, the data were tabulated, and each outcome was grouped into one of the following six outcome domains, some of which were further divided into subdomains where appropriate: disease activity, physical consequence of disease, functional status, social outcomes and quality of life, side effects of therapy and health resource utilisation. These domains had been identified in the systematic review described in Chapter 2.

To assess how the selection of outcomes has changed over time, the period 1988 to 2007 was divided into sixteen separate epochs, each lasting five years. In each epoch, the proportion of studies measuring each outcome domain was calculated, and the results presented as a moving window. This was only done using studies that had exclusively involved children.

3.4 Results

3.4.1 Results of the search

The search yielded 1668 potentially eligible reports. Of these, 1256 were excluded by reading the abstract. The remaining 412 were retrieved in full, and 203/412 were subsequently excluded. In total, 209 eligible reports, of 159 RCTs, were included in the review. The review flowchart is shown in Figure 2.
Figure 2 Flowchart of the review: identification of RCTs, assessing ICS in children with asthma, published between January 1988 and December 2007

Abstracts screened 1668

Excluded by reading abstract 1256

Full text reports retrieved 412

Full text report excluded 203
  - 107 did not include children
  - 30 were crossover studies
  - 28 were not studies of ICS
  - 15 were not RCTs
  - 12 had treatment period <4 weeks
  - 11 not in children with asthma

Report included 209
(159 separate RCTs)
3.4.2 Description of included studies

Of the 159 studies included in this review, 115 exclusively included children, and 44 included children and adults. Within the group of studies that included only children, 25/115 (21%) included children younger than four years of age. In the studies of adults and children, 42/44 (95%) included children aged between 12 and 18 years of age, but not younger than 12, and 2/44 (5%) included children between 5 and 18 years of age, but not younger than 5 years.

Of the 159 studies, 83 (52%) included a comparison between ICS groups (either different doses, modes of delivery, or types of ICS, eg budesonide vs beclomethasone), 63 (40%) included a comparison with placebo, and 54 (34%) included a comparison with another drug.

Masking of interventions was deemed adequate in 121/159 studies (76%), inadequate in 33/159 (21%), and unclear in 5/159 (3%). Of the 33 studies that were classed as inadequately masked, 18 compared ICS with another drug, 8 compared ICS administered by different devices, 3 compared one ICS to another, 3 compared ICS administered using different dosing schedules, and 1 compared ICS with no treatment. Subjective outcomes that could have been affected by the lack of blinding were measured in 29/33 of these studies (29/33 measured symptoms, 8/33 measured quality of life, and 6/33 measured functional status).

The characteristics of included studies are summarised in Table 3.
### Table 3 Characteristics of included studies

<table>
<thead>
<tr>
<th>Study characteristic</th>
<th>Category</th>
<th>Number (%) of RCTs that included only children (n=115)</th>
<th>Number (%) of RCTs which included children and adults (n=44)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date of publication</td>
<td>January 1988 to December 1992</td>
<td>10 (9)</td>
<td>1 (2)</td>
</tr>
<tr>
<td></td>
<td>January 1993 to December 1997</td>
<td>19 (17)</td>
<td>5 (11)</td>
</tr>
<tr>
<td></td>
<td>January 1998 to December 2002</td>
<td>48 (42)</td>
<td>22 (50)</td>
</tr>
<tr>
<td></td>
<td>January 2003 to December 2007</td>
<td>38 (32)</td>
<td>16 (37)</td>
</tr>
<tr>
<td>Length of treatment period</td>
<td>1 to 3 months</td>
<td>30 (26)</td>
<td>5 (11)</td>
</tr>
<tr>
<td></td>
<td>3 to &lt;6 months</td>
<td>46 (40)</td>
<td>30 (69)</td>
</tr>
<tr>
<td></td>
<td>6 to &lt;12 months</td>
<td>13 (11)</td>
<td>4 (9)</td>
</tr>
<tr>
<td></td>
<td>12 months or longer</td>
<td>24 (21)</td>
<td>5 (11)</td>
</tr>
<tr>
<td></td>
<td>Unclear</td>
<td>2 (2)</td>
<td>0</td>
</tr>
<tr>
<td>Age groups of children included</td>
<td>&lt;4 years only</td>
<td>17 (15)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>&lt;4 and 4 to &lt;12 years</td>
<td>7 (6)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>&lt;4 and 4 to&lt;12 and 12 to 18 years</td>
<td>1 (1)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>4 to &lt;12 years only</td>
<td>36 (31)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>4 to &lt;12 years and 12 to 18 years</td>
<td>53 (46)</td>
<td>2 (5)</td>
</tr>
<tr>
<td></td>
<td>12 to 18 years only</td>
<td>1 (1)</td>
<td>42 (95)</td>
</tr>
<tr>
<td>Number of centres</td>
<td>Multicentre</td>
<td>75 (65)</td>
<td>42 (95)</td>
</tr>
<tr>
<td></td>
<td>Single centre</td>
<td>39 (34)</td>
<td>2 (5)</td>
</tr>
<tr>
<td></td>
<td>Unclear</td>
<td>1 (1)</td>
<td>0</td>
</tr>
<tr>
<td>Source of funding</td>
<td>Industry</td>
<td>85 (74)</td>
<td>42 (95)</td>
</tr>
<tr>
<td></td>
<td>Public funding bodies</td>
<td>30 (26)</td>
<td>2 (5)</td>
</tr>
<tr>
<td></td>
<td>ICS vs Other drug</td>
<td>26 (23)</td>
<td>12 (27)</td>
</tr>
<tr>
<td></td>
<td>ICS vs placebo</td>
<td>25 (22)</td>
<td>4 (9)</td>
</tr>
<tr>
<td></td>
<td>ICS 1 vs ICS 2 a</td>
<td>14 (12)</td>
<td>3 (7)</td>
</tr>
<tr>
<td></td>
<td>ICS vs same ICS (other device)</td>
<td>12 (10)</td>
<td>2 (5)</td>
</tr>
<tr>
<td></td>
<td>ICS vs same ICS (other dose)</td>
<td>9 (8)</td>
<td>3 (7)</td>
</tr>
<tr>
<td></td>
<td>ICS vs no treatment</td>
<td>1 (1)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>ICS 1 vs ICS 2 vs Placebo a</td>
<td>16 (14)</td>
<td>10 (23)</td>
</tr>
<tr>
<td></td>
<td>ICS 1 vs ICS 2 vs Other drug a</td>
<td>7 (6)</td>
<td>1 (2)</td>
</tr>
<tr>
<td></td>
<td>ICS vs Other drug vs placebo</td>
<td>4 (3)</td>
<td>4 (9)</td>
</tr>
<tr>
<td></td>
<td>ICS1 vs ICS 1 (other dose and device)</td>
<td>1 (1)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>ICS 1 vs ICS 1 (other dose) vs ICS 2</td>
<td>1 (1)</td>
<td>0</td>
</tr>
</tbody>
</table>

Footnote: a: ICS 1 vs ICS 2 refers to comparison of two different classes of ICS
3.4.3 Outcome domains measured in the studies

Disease activity was measured in 157/159 (99%) studies, adverse effects of therapy in 135/159 (85%), functional status in 25/159 (16%), quality of life in 21/159 (13%), and health resource utilisation in 17/159 (11%).

No studies measured the effects of ICS on long-term physical consequences of asthma, although two studies measured post-bronchodilator FEV1, as a percentage of the predicted value, to assess ‘lung growth’. In one of these studies, children, aged between 5 and 12 years, were randomised to receive inhaled budesonide, nedocromil sodium, or placebo, for a period of four to six years, and FEV1 was measured as the primary outcome (Szeffler et al 2000). In the other study, patients aged between 5 and 66 years were randomised to treatment with inhaled budesonide or placebo for three years, and FEV1 was measured as a secondary outcome (Pauwels et al 2003).

Similar outcome domains were represented in the 115 studies that included only children and the 44 that included children and adults.

3.4.4 Outcomes that were measured within the domains

Subdomains and outcomes varied across the studies. There was a wide variety of outcomes within individual domains. This was greatest for the disease activity domain, which was divided into five subdomains (clinical measures, physiological tests of lung function, global measures, bronchial
responsiveness to a challenge agent or exercise and markers of inflammation). The frequency with which domains, subdomains and outcomes were selected, in 115 studies that exclusively involved children, is shown in Table 4.
Table 4 Frequency with which outcomes were reported in the 115 RCTs which involved only children

<table>
<thead>
<tr>
<th>Domain</th>
<th>Subdomain 1</th>
<th>Subdomain 2</th>
<th>Outcome</th>
<th>Number (%) of studies in which measured (primary or secondary outcome) n=115</th>
<th>Number (%) of studies in which measured (primary outcome) n=84 a</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Symptom severity</td>
<td>90 (77)</td>
<td>10 (12)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Symptom frequency</td>
<td>55 (47)</td>
<td>5 (6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Use of rescue therapy</td>
<td>90 (77)</td>
<td>2 (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Exacerbation frequency</td>
<td>35 (30)</td>
<td>4 (5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Time to exacerbation</td>
<td>10 (9)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FEV1</td>
<td>80 (70)</td>
<td>16 (19)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FVC</td>
<td>31 (26)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mid expiratory flow</td>
<td>23 (20)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FEV1:FVC</td>
<td>6 (5)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FEV1 reversibility</td>
<td>9 (8)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>PEFR</td>
<td>85 (73)</td>
<td>26 (31)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Diurnal variability</td>
<td>13 (11)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Day-to-day variability</td>
<td>5 (4)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Spirometry</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Resistance/conductance</td>
<td>5 (4)</td>
<td>2 (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Physician-rated</td>
<td>8 (7)</td>
<td>1 (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Parent/patient – rated</td>
<td>14 (12)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>‘Treatment failure’</td>
<td>13 (11)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>‘Treatment success’</td>
<td>3 (3)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Bronchial responsiveness to a challenge agent n=29</td>
<td>26 (22)</td>
<td>6 (7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Exhaled nitric oxide</td>
<td>5 (4)</td>
<td>1 (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Leukotriene b interleukin c</td>
<td>4 (3)</td>
<td>1 (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Eosinophils d IgE e</td>
<td>18 (15)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>HRU</td>
<td>15 (13)</td>
<td>1 (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Unscheduled HRU</td>
<td>15 (13)</td>
<td>1 (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Functional status</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Effect of asthma on ADL</td>
<td>10 (9)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>School attendance</td>
<td>15 (13)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>QoL/ family outcomes</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Child’s QoL</td>
<td>9 (8)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Caregiver QoL</td>
<td>5 (4)</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Caregiver functional status</td>
<td>8 (7)</td>
<td>0</td>
</tr>
<tr>
<td>Adverse Effects of therapy</td>
<td>Subdomain 1</td>
<td>Subdomain 2</td>
<td>Outcome</td>
<td>Number (%) of studies in which measured (primary or secondary outcome) n=115</td>
<td>Number (%) of studies in which measured (primary outcome) n=84 a</td>
</tr>
<tr>
<td>---------------------------</td>
<td>-------------</td>
<td>-------------</td>
<td>---------</td>
<td>---------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td>Routinely monitored AE n=82</td>
<td>Patient/parent-reported</td>
<td></td>
<td></td>
<td>80 (70)</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Routine laboratory AE</td>
<td></td>
<td></td>
<td></td>
<td>32 (27)</td>
<td>0</td>
</tr>
<tr>
<td>Oropharyngeal infection</td>
<td></td>
<td></td>
<td></td>
<td>28 (24)</td>
<td>0</td>
</tr>
<tr>
<td>Ophthalmological events</td>
<td></td>
<td></td>
<td></td>
<td>7 (6)</td>
<td>0</td>
</tr>
<tr>
<td>H-P-A axis n=52</td>
<td>Urine/serum cortisol</td>
<td></td>
<td></td>
<td>52 (44)</td>
<td>0</td>
</tr>
<tr>
<td>ACTH stimulation</td>
<td></td>
<td></td>
<td></td>
<td>17 (15)</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Growth n=41</td>
<td>Growth</td>
<td></td>
<td></td>
<td>41 (35)</td>
<td>9 (11)</td>
</tr>
<tr>
<td>Lower leg growth</td>
<td></td>
<td></td>
<td></td>
<td>1 (1)</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Effects of ICS on bone n=15</td>
<td>Markers of bone turnover</td>
<td></td>
<td></td>
<td>11 (9)</td>
<td>0</td>
</tr>
<tr>
<td>Measures of bone density</td>
<td></td>
<td></td>
<td></td>
<td>8 (7)</td>
<td>2 (2)</td>
</tr>
</tbody>
</table>

**Abbreviations used in Table 4:**
ACTH= adrenocorticotropic hormone; ADL=Activities of Daily Living; AE= adverse events; BHR= bronchial hyperresponsiveness; FEV1= forced expiratory flow in one second; FVC= forced vital capacity; H-P-A=Hypothalamic-pituitary-adrenal; HRU= health resource utilisation; ICS=inhaled corticosteroids; IgE= Immunoglobulin E; PEFR= peak expiratory flow rate; QoL= Quality of life

**Footnotes to Table 4:**
a: Of 84 studies that specified primary outcomes, 5 specified co-primary outcomes, and hence the total number of primary outcomes measured is 89.
b: Leukotriene LTC4 in serum and nasal secretions and leukotriene LTE4 in urine
c: Interleukins in serum and sputum
d: Eosinophils in serum and sputum, and Eosinophil Cationic Protein in serum and urine
e: IgE in serum
3.4.5 Primary outcomes

It was possible to determine the primary outcome in 84/115 (73%) of the studies that included only children. In 64/84, the primary outcome was clearly stated by the authors, and in 20/84 it was inferred from the outcome used to calculate the sample size. Five studies each selected two co-primary outcomes.

Of the 84 studies that specified a primary outcome, 74 (88%) selected primary outcomes that measured disease activity. A total of 17 different primary outcomes were selected, of which physiological measures of airway obstruction, mainly PEFR (26 studies) and FEV1 (16 studies), were the most frequent. None of the primary outcomes addressed the functional status or quality of life domains.

In RCTs that included children and adults, it was possible to determine the primary outcome in 39/44 (89%) studies. In 34 of these, the primary outcome was clearly stated by the authors, and in the remaining 5 it was inferred from the outcome used to calculate the required sample size. In 38/39 (97%) of studies, the primary outcome was a measure of disease activity. The most frequently measured primary outcome was FEV1 (28 studies).
3.4.6 Outcomes measured in studies funded by the pharmaceutical industry

The frequency with which most domains were measured was similar between the 127 studies sponsored by the pharmaceutical industry and the 32 publicly funded studies. The main difference was that adverse effects of therapy were reported in a higher proportion of studies sponsored by the pharmaceutical industry (118/127, 93%) compared to studies funded from other sources (17/32, 53%).

3.4.7 Trends over time in the selection of outcome domains

The trend over the period January 1988 to December 2007 in the selection of outcome domains in RCTs that only included children is shown in Figure 3. Disease activity and adverse effects of therapy have consistently remained frequently measured outcome domains. Since the 1992–1996 epoch the proportion of studies measuring functional status, for example by assessing school absence due to asthma, has decreased from 40% to 10%, and those measuring quality of life have increased from 10% to 25%.
Figure 3 Trends in the measurement of outcome domains between 1988 and 2007

Legend for Figure 3: The figure shows trends in the measurement of outcome domains in clinical trials of inhaled corticosteroid for children with asthma published between 1988 and 2008. Data are presented as a moving window. Each point along the horizontal axis represents the midpoint of a five year epoch. In each epoch the proportion of studies measuring each individual domain is shown.
3.5 Discussion

We found that RCTs in children with asthma almost always assess the effects of therapies on short-term disease activity, but not the effects on long-term progression of disease. Quality of life and functional status are measured infrequently. While there were certain similarities between studies, particularly in the selection of primary outcomes that measure disease activity, the selection and measurement of outcomes showed wide variability.

The pharmaceutical industry funded 80% of the RCTS we identified, and so it is not surprising that outcomes in the disease activity domain have been measured, as primary or secondary outcomes, as this reflects the requirements of the FDA (FDA 1994; FDA 2007) and EMA (EMA 1993; EMA 2002; EMA 2004; EMA 2007). These authorities recommend, for the purpose of drug licensing and marketing authorisation, that clinical trials of preventative therapies for children with asthma should measure, as primary outcomes, physiological tests of pulmonary function, and clinical measures such as symptom scores. Other measures of short-term disease activity, such as use of rescue medication, rate of exacerbations, and bronchial hyper-responsiveness are suggested as important outcomes. QoL and exercise tolerance are mentioned as additional outcomes that may provide useful information, but are not explicitly recommended.

Given that the aim of an RCT is to evaluate the safety and efficacy of interventions, and provide some assessment of whether it does more good
than harm, it is disappointing to note that QoL and family outcomes are only measured in 20% of RCTs in the last five years worth of studies included in the review, and that measures of functional status, such as school attendance, are reported in less than 10%. We identified no clinical trials in which the primary outcome measured these domains. In studies that have investigated which outcomes are clinically relevant to patients with other chronic conditions, such as long-term pain (Dworkin et al 2005), fibromyalgia (Mease et al 2008), and rheumatoid arthritis (Hewlett et al 2005), measures of functional status and quality of life were identified as being of great importance, and it is likely that this is also the case in children with asthma. Functional status overlaps with quality of life and disease activity. However, we feel that measures of functional status are important, distinct, markers of how asthma affects children. In clinical trials in adults with chronic illnesses, absence from work is an important outcome, and we feel that an appropriate childhood equivalent would include measures of school attendance and other activities of daily living.

It is particularly important in trials of children to assess the impact of treatments in the long term. Very few studies have attempted to study the impact of ICS on modifying or affecting the physical consequences of asthma. Two studies, the Childhood Asthma Management Program (Childhood asthma management program research group 2000), and the Inhaled Steroid Treatment as Regular Therapy in Early Asthma (START) (Pauwels et al 2003), indicated that they wished to investigate this effect. Both measured primary outcomes that we considered related to disease activity, although the CAMP study stated that their primary outcome
(FEV1, %predicted) was a measure of ‘lung growth’. Although it is commendable that these trialists attempted to assess long-term benefits of ICS, investigators acknowledge the difficulties in assessing the impact of disease, or therapy, on ‘lung growth’ in children with asthma. It is unlikely that, in a clinical trial, a single measure will provide the best primary outcome, and more methodological research is needed to identify whether longitudinal outcomes, such as serial measurement of lung function, would be a more appropriate way of assessing lung growth. Although 24 of the studies that exclusively included children had a treatment period lasting longer than one year, only three measured outcomes after the end of the treatment period (Szeffler et al 2000; Visser et al 2004; Merkus et al 2004). The other 21 studies represent missed opportunities to investigate the long term effects of ICS on progression of asthma in children, and this question should be addressed in future clinical trials.

The long-term safety of treatments for asthma has recently been identified as being of particular importance to patients and clinicians (James Lind Alliance 2007). Although 83% of studies that we identified, that only included children, assessed the safety of ICS, the quality with which long-term systemic side effects of ICS were measured was variable. Of the 24 studies that lasted longer than a year, 21 measured effects on growth. None of the studies measured the final adult height attained, despite the fact that this is of most interest to children and parents. Only 14 studies measured the effect of ICS, when administered for longer than one year, on hypothalamo-pituitary-adrenal function, despite the serious and potentially fatal consequences of this adverse reaction (Paton et al 2006). We suggest
that serious systemic side effects should be monitored in all clinical trials of ICS in children with asthma, so that both the benefits and risks of these drugs can be appropriately evaluated.

As well as the fact that measurement of long-term efficacy and safety outcomes is not a requirement of drug regulatory authorities, there are other reasons why they may not have been measured in the studies we identified. Diagnostic and technical problems associated with measurement of lung growth, financial cost, and problems with patient attrition, also hinder the conduct of long-term studies in children with asthma. Agreement that long-term outcomes are important, amongst clinicians, patients, and researchers, could promote the conduct of such studies. There is also a need for research to identify the most appropriate long-term outcomes, and the ways in which they should be measured.

**Robustness of the review**

This review represents a thorough analysis of RCTs spanning twenty years, and this enabled us to thoroughly evaluate the outcome domains that were measured, and whether this changed over time. Even though we have reviewed RCTs assessing one aspect of the treatment of childhood asthma, it is likely that our findings would be similar if we were to conduct a similar review of, for example, clinical trials of LTRA or LABA. The studies we identified are comparable in terms of the population they
include and the interventions that they compare, and so we feel that our finding of heterogeneity of outcome selection between studies is valid.

We have reviewed outcomes that have been reported rather than those which were actually measured. ORB in published RCT reports is common (Chan, & Altman 2005; Dwan et al 2008), and in order to have evaluated exactly which outcomes had been measured it may have been more accurate to assess trial protocols. Although ORB may lead to an underestimation of the frequency with which some outcomes were actually measured, it is unlikely that it would affect heterogeneity between studies.

We chose to identify RCTs using the Cochrane Database of Controlled Trials because this database includes publications that are not held in Medline, and everything held within it has been categorised as being a controlled trial. In order to search for eligible RCTs even more thoroughly, if time constraints were not an issue, we could also have searched Medline. The main reason for doing this is that Cochrane databases are updated quarterly, whereas Medline is updated weekly (Glanville et al 2006). It is unlikely, however, that many studies were missed by not searching Medline.

3.6 Summary

We have shown that outcomes in RCTs in children with asthma are focussed on short-term disease activity, and other domains are largely
overlooked. Studies are not uniform with regard to the selection, measurement, and reporting of outcomes. This can make it difficult to design, interpret, and meta-analyse clinical trials. One solution is a core set of outcomes that should be measured and reported, as a minimum, in all clinical trials of a specific condition. Core sets were first designed by the OMERACT group, which utilises structured consensus techniques, including the Delphi technique, amongst a diverse group of stakeholders and consumers. Our findings would suggest that a similar initiative in childhood asthma, with separate consideration of pre-school and older children, would make an important contribution to improving clinical research in this very prevalent, chronic disease. In Chapter 2, we showed that such a core set had not already been developed.

In order to develop this core outcome set, it is important to review, in detail, other studies which have used structured consensus techniques to determine which outcomes to measure in clinical trials. In Chapter 2 we identified that the Delphi technique was the most widely used structured method for reaching consensus, whereas NGT could only be used amongst a smaller panel, and so, by itself, was not used as the only technique for developing a core outcome set. A systematic review of studies, which have used the Delphi technique to identify which outcomes to measure in clinical trials, in adults or children, is described in Chapter 4.
Chapter 4

A systematic review of studies that have used the Delphi process as a method for determining which outcomes to measure in clinical trials

4.1 Background

The standardisation of outcomes for clinical trials has been proposed as a solution to the problems of inappropriate and non-uniform outcome selection (Miller et al 2001; Clarke 2007) and reporting bias (Giannini et al 1997; Williamson et al 2005; Kirkham et al 2010). The most notable work relating to outcome standardisation has been conducted by the OMERACT collaboration, which advocates the use of core outcome sets, designed using consensus techniques, which are measured and reported in clinical trials in rheumatology (Tugwell et al 2007). However, such initiatives are uncommon. In some specialties, such as paediatrics, the number of conditions covered is low and the quality of existing studies variable.

One method for reaching consensus around which outcomes to measure is the Delphi technique, which comprises sequential questionnaires, answered anonymously, by a panel of participants, who each have relevant expertise. After each questionnaire, the group response is fed back to participants (Dalkey, & Helmer 1963). With regard to the overall validity of the final
consensus, this approach has advantages over open discussions. Participants in a Delphi study do not interact directly with each other, so situations where the group is dominated by the views of certain individuals can be avoided. When participants consider whether to change their opinion or stick to their original answers, after seeing the group response, this decision is not affected by the desire to be seen to agree with senior, overly vocal, or domineering individuals. Improvements in global communication have made it feasible to use the Delphi technique to involve geographically distant participants, in larger numbers than are traditionally used in studies employing face-to-face discussion, and so it is also increasingly being used to reach consensus around other topics in medicine, such as education (Alahlafi, & Burge 2005), clinical guidelines (Morita et al 2005), and prioritisation of research topics (Kellum et al 2008).

There is little guidance for researchers who wish to use the Delphi technique, even though aspects of its methodology can be interpreted in a variety of ways. Most published work has provided guidance based on authors’ experiences, rather than empirical research or theoretical justification for the methodological decisions made. One systematic review describes a variety of consensus techniques in the context of designing clinical guidelines (Murphy et al 1998). The authors highlighted important methodological decisions that may affect the overall quality of the final consensus, including the types of participants involved, the questions they are asked, the information they receive to inform their answers, the manner of the interaction between them, and the way in which consensus is agreed.
To our knowledge, there is no guidance about how to conduct or report studies that use the Delphi technique to determine which outcomes or domains to measure in clinical trials or systematic reviews. The work described in this chapter is a systematic review of studies that have used the Delphi technique for this purpose.

### 4.2 Aims

1) To identify variations in the methods and reporting of studies that use the Delphi technique to determine which outcomes, or domains, to measure in clinical trials or systematic reviews.

2) To discuss recommendations relating to the design and reporting of future studies, based on the findings of our review, for people who wish to use the Delphi technique for this purpose.

### 4.3 Methods

#### 4.3.1 Inclusion and exclusion criteria

Studies were eligible for inclusion if they used the Delphi technique to determine which outcomes or domains to measure in clinical trials or systematic reviews. We included both studies in which the final consensus was determined using the Delphi technique, and also studies that used the
Delphi technique to identify the opinion of a group of participants, but not in order to reach a final consensus about which outcomes to measure. This latter category includes studies in which the results of the Delphi technique were used to inform participants in subsequent meetings to agree which outcomes or domains to measure in clinical trials or systematic reviews.

We felt it was appropriate to combine studies relating to outcomes with those relating to outcome domains, because we expected the methods used in both these types of studies to be sufficiently similar.

Studies that did not specifically state that the outcomes selected could be used in clinical trials or systematic reviews were excluded. For example, studies identifying outcomes to evaluate the quality of care given by healthcare providers, or to guide the management of individual patients in clinical practice, were not included, unless the authors stated that participants considered their use in clinical research.

### 4.3.2 Identification of studies, and the search strategy

Medline was searched in January 2010 (from 1950 to January 2010). The process of designing a search strategy is described in Section 2.3.3. The following search strategy was used in this review:

“((outcome$ OR endpoint$ OR end point$ OR variable$ OR domain$) AND (Delphi OR Delphi method OR Delphi technique$ OR sequential questionnaire$)).mp.”.
4.3.3 Data extraction

The following methodological aspects were identified from each study report:

1. Who was involved in the Delphi process, in terms of the number of clinicians, patients and other participants, and their geographical location

2. The types of questions participants were asked

3. Whether the study was completely anonymised (participants’ identities and answers were hidden from the group) or quasi-anonymised (identities of participants were disclosed to the group, but individual answers were not)

4. Whether non-responders in earlier rounds were included or excluded from subsequent rounds

5. The definition of consensus used by the authors.

We contacted the corresponding author by email to obtain information that was not available or clear in the study report.

4.3.4 Assessment of reporting quality

We assessed in the study report whether the following methodological aspects were described:

1. The number of participants invited to each round

2. The types of participants involved, and the number of each type
3. How participants were initially identified as being eligible for the study

4. The medium used to conduct the Delphi study (e.g. postal, email, internet)

5. What information participants received before the first round

6. How outcomes included in the first round of questions were identified (i.e. were they identified before the study, or did participants suggest outcomes that should be considered by the group)

7. Whether the study was completely anonymised or quasi-anonymised

8. What questions were asked in each round

9. What feedback was provided to participants after each round

10. How the authors determined, for each outcome, whether consensus had been reached as to whether it should be measured.

If any of these were not described to the level of detail that would permit another researcher to reproduce the methodology, we classed the reporting of the methodological aspect as unclear. We also identified whether the authors explained these methodological choices.

With regard to the reporting of results, we assessed whether the authors had described the following:

1. The number of respondents to each round

2. The proportion of participants who completed every round
3. The results for each outcome in each round

4. A measure of group response and distribution for each outcome in the final round

5. A list of all outcomes that the group decided should be measured in clinical research studies.

4.3.4 Data analysis and presentation of results

For synthesis of data the studies were described narratively. Consistent with the nature of the data, the results are presented in textual format.

4.4 Results

4.4.1 Identification of studies

Of twenty studies for which the full text study report was retrieved, five were excluded because they aimed to identify outcomes for use in clinical practice, and the authors did not state whether the participants considered their use in clinical research studies (Brunner et al 2008; Jones et al 2000; van Hulst et al 2009; Weigl et al 2004; Radtke et al 2009). Fifteen studies were included in the review. The review flowchart is shown in Figure 4.
Figure 4 Flowchart of the review: Identification of studies that used the Delphi process to determine which outcomes or domains to measure in clinical trials or systematic reviews.
Eight of the included studies developed core outcome sets for rheumatological conditions, of which five were conducted by the OMERACT group (relating to gout (Taylor et al 2008), fibromyalgia (Mease et al 2008), psoriatic arthropathy (Taylor 2005), and systemic sclerosis (Khanna et al 2008), and associated pulmonary hypertension (Distler et al 2008)) and three were conducted by other collaborations (relating to idiopathic inflammatory myopathy (Miller et al 2001), juvenile systemic lupus erythematosus (Ruperto et al 2003), and ankylosing spondylitis (Zochling et al 2008)). Others developed core outcome sets for pain in children (McGrath et al 2008), degenerative ataxia (Serrano-Aguilar et al 2009), gastro oesophageal reflux disease (Dent et al 2008), infantile spasms (Lux, & Osborne 2004), maternity care (Devane et al 2007), multiple sclerosis (Khan, & Pallant 2007), and thyroid eye disease (Douglas et al 2009). The methods and results of one study were described across two publications, of which the one that describes the study most comprehensively (Miller et al 2001) is included in this review, and the other is referenced as an additional article (Oddis 2005).

One study was conducted in two distinct stages, both of which were reported in separate publications. The most recent publication describes both stages of the study (Mease et al 2008), and is included in this review, while the earlier publication is referenced as an additional article (Mease et al 2005).

Eight studies used the Delphi technique as the main method of reaching consensus about which outcomes to measure in clinical research studies. Four of these related to outcomes in rheumatological conditions (of which three were conducted by the OMERACT collaboration (Distler et al 2008; Khanna
et al 2008; Taylor et al 2008) and one by another group (Zochling et al 2008), two related to neurological conditions (Khan, & Pallant 2007; Lux, & Osborne 2004), one related to gastroenterology (Dent et al 2008) and one to maternity care (Devane et al 2007). The remaining seven studies used the results of a Delphi process either to inform people involved in subsequent consensus studies (Miller et al 2001; Douglas et al 2009; McGrath et al 2008; Mease et al 2008; Ruperto et al 2003; Taylor 2005), or to help design a specific systematic review (Serrano-Aguilar et al 2009).

4.4.2 Reporting quality

The reporting quality of the studies is summarised in Table 5.
Table 5 Summary of the reporting quality of the studies which used the Delphi technique to determine which outcomes or domains to measure in clinical trials or systematic reviews

<table>
<thead>
<tr>
<th></th>
<th>Studies in which clearly reported</th>
<th>Studies in which not clearly reported</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Size and composition of the panel</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of participants</td>
<td>15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Types of participants (eg clinicians, patients)</td>
<td>15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Proportion of each type of participant</td>
<td>15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>How participants were identified/sampled</td>
<td>14</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><strong>Methodology of the Delphi process</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Administration of questionnaires (eg postal)</td>
<td>15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>How items were generated for first questionnaire</td>
<td>14</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>What was asked in each round</td>
<td>15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Information provided to participants before the first round</td>
<td>6</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>How the overall group response was fed back to participants</td>
<td>8</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>Level of anonymity (total or quasi-anonymity)</td>
<td>4</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>A priori definition of ‘consensus’ about whether an outcome should be measured</td>
<td>7</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Were non-responders invited to subsequent rounds</td>
<td>10</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td><strong>Results</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of respondents to each round</td>
<td>14</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Number who completed every round</td>
<td>11</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Results for each outcome in each round</td>
<td>0</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>Group response for each outcome (final round)</td>
<td>8</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>Distribution of response for each outcome in the final round</td>
<td>7</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>List of all outcomes that participants agreed should be measured</td>
<td>8</td>
<td>0</td>
<td>7</td>
</tr>
</tbody>
</table>

**Footnotes to Table 5:**

a: Reaching a final consensus was not the aim of the Delphi process, so a definition of consensus was not given

b: All participants responded to each round so no discussion was made regarding non-responders
**Reporting of methods**

The types of participants involved, and the size and composition of the group, were adequately described in all studies. All but one study described the way in which, at the start of the Delphi process, an initial list of outcomes was identified, for consideration by the panel. All studies described the questions that were asked, during the rounds of the Delphi process, to determine which of these should be measured in clinical trials. All studies described, when applicable, the pre-determined definition of ‘consensus’ about whether an outcome should be measured in clinical trials.

Important methodological aspects that were generally less well reported were the information provided to participants at the start of the Delphi process (clearly reported in 6/15 studies), the information that was fed back to participants after each round (clearly reported in 8/15 studies), and the level of anonymity in the study (clearly reported in 4/15 studies).

**Reporting of results**

An assessment of response rate could be made in 14/15 studies, which reported, for each round, the number of invited participants who responded. It was possible to make an assessment of attrition rates in 11/15 studies, which reported the proportion of first round respondents who also completed the final round. Of these, six studies reported the proportion of participants who completed every round in the Delphi process, from start to finish.
Although all studies provided a list of outcomes that participants felt should be measured in clinical research studies, only eight reports presented a measure of the group opinion for each outcome listed in the final round, and seven of these also reported the distribution of scores for each item. No study reported the results, in each round, for every outcome that was considered by the group.

**Explanation for methods adopted in the study report, or discussion of the impact of methodological decisions on the final results**

The composition of the groups was discussed in ten studies, but only six discussed the implications of the size of their group on the external validity of their results. Two studies discussed why they chose to either identify outcomes in a certain way at the start of the Delphi process, or to let participants determine the initial list of outcomes. Five studies explained the methods used to determine which of these initial outcomes should be measured in clinical trials. Five studies discussed why they defined consensus in a certain way. Two studies discussed the implications of presenting results to participants in a certain way.
4.4.3 Methodological variation between the studies

Composition of the group

The group size varied from 13 (Dent et al 2008) to 222 (Ruperto et al 2003). Generally, studies conducted through clinical or research networks involved more participants.

Clinicians were included in all but one study, which only involved patients (Serrano-Aguilar et al 2009). Eight studies were conducted through clinical trial networks (Miller et al 2001; Khanna et al 2008; Lux, & Osborne 2004; Mease et al 2008; Ruperto et al 2003; Taylor 2005; Taylor et al 2008; Zochling et al 2008), four only involved clinicians who had published research in the relevant field (Dent et al 2008; Distler et al 2008; Lux, & Osborne 2004; McGrath et al 2008), and three involved both clinicians and researchers (Devane et al 2007; Douglas et al 2009; Khan, & Pallant 2007).

Four groups involved patients or families (Devane et al 2007; Serrano-Aguilar et al 2009; Zochling et al 2008; Mease et al 2005). Some groups involved other types of participants, including health service managers (Devane et al 2007), pharmaceutical industry employees (McGrath et al 2008; Taylor et al 2008), and drug regulatory agency representatives (McGrath et al 2008).

Five studies involved different types of participants, of which four used a single panel comprising a mix of the groups (Devane et al 2007; McGrath et al 2008; Taylor et al 2008; Zochling et al 2008). In the other study, relating to a core set of outcomes for fibromyalgia, clinicians and patients completed
two separate Delphi studies, which were used to inform discussions at a subsequent consensus meeting (Mease et al 2008; Zochling et al 2008).

The size and composition of the groups is described in Table 6.
Table 6 Size and composition of the groups involved in Delphi processes to determine which outcomes to measure in clinical trials

<table>
<thead>
<tr>
<th>Study</th>
<th>Method by which sample of identified</th>
<th>Response to Round 1 (number)</th>
<th>Number (%) who were health care providers</th>
<th>Number (%) who were patients</th>
<th>Other people involved in the study</th>
<th>Research experience of the panel</th>
<th>Number of countries represented (Continents)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dent</td>
<td>People known to facilitator</td>
<td>13</td>
<td>12 (92)</td>
<td>0</td>
<td>Clinical trial methodologist n=1 (8)</td>
<td>1 member (non-clinician) was an expert in RCT methods; 9 (North America, Europe, Asia, Australasia)</td>
<td></td>
</tr>
<tr>
<td>Devane</td>
<td>Health professional network; patient groups</td>
<td>218</td>
<td>147 (68)</td>
<td>24 (11)</td>
<td>Health service managers n=14(6); epidemiologists n=9(4); ‘other’ 24 (11)</td>
<td>78/218 (36%) self-identified as researcher</td>
<td>13 (North America, South America, Europe, Asia, Australia, Asia)</td>
</tr>
<tr>
<td>Distler</td>
<td>Clinical trial network</td>
<td>69</td>
<td>69 (100)</td>
<td>0</td>
<td>0</td>
<td>All had published clinical research</td>
<td>Unclear (North America, Europe, Australasia, Asia)</td>
</tr>
<tr>
<td>Douglas</td>
<td>Health professional network</td>
<td>84</td>
<td>84 (100)</td>
<td>0</td>
<td>0</td>
<td>Unclear</td>
<td>14 (North America, Europe, Australasia, Asia)</td>
</tr>
<tr>
<td>Khan</td>
<td>Local professional</td>
<td>23</td>
<td>23 (100)</td>
<td>0</td>
<td>0</td>
<td>Unclear</td>
<td>1 (Australia)</td>
</tr>
<tr>
<td>Khanna</td>
<td>Clinical trial network</td>
<td>62</td>
<td>62 (100)</td>
<td>0</td>
<td>0</td>
<td>All were members of a clinical trial network.</td>
<td>Unclear (North America, South America, Europe, Asia)</td>
</tr>
<tr>
<td>Lux</td>
<td>Published researchers</td>
<td>31</td>
<td>31 (100)</td>
<td>0</td>
<td>0</td>
<td>All participants had presented/published clinical research</td>
<td>15 (North America, South America, Europe, Asia)</td>
</tr>
<tr>
<td>Mease</td>
<td>Clinical trial network; Local patients</td>
<td>96</td>
<td>23 (24) b</td>
<td>73 (76)</td>
<td>0</td>
<td>All clinicians were members of a clinical trial network.</td>
<td>1 (North America) Nb consensus meeting at OMERACT, after the clinician Delphi, was multinational</td>
</tr>
<tr>
<td>McGrath</td>
<td>People known to facilitator</td>
<td>26</td>
<td>17 (65)</td>
<td>0</td>
<td>FDA/NIH n=5 (19); industry n=4 (16)</td>
<td>All doctors were researchers</td>
<td>4 (North America, Europe)</td>
</tr>
<tr>
<td>Miller</td>
<td>Clinical trial network</td>
<td>70</td>
<td>70 (100)</td>
<td>0</td>
<td>0</td>
<td>All were investigators in a clinical trial network</td>
<td>14 (North America, South America, Europe, Asia)</td>
</tr>
<tr>
<td>Study</td>
<td>Method by which sample of identified</td>
<td>Response to Round 1 (number)</td>
<td>Number (%) who were health care providers</td>
<td>Number (%) who were patients</td>
<td>Other people involved in the study</td>
<td>Research experience of the panel</td>
<td>Number of countries represented (Continents)</td>
</tr>
<tr>
<td>-----------------------</td>
<td>--------------------------------------</td>
<td>------------------------------</td>
<td>------------------------------------------</td>
<td>-------------------------------</td>
<td>----------------------------------</td>
<td>---------------------------------</td>
<td>--------------------------------------------</td>
</tr>
<tr>
<td>Ruperto</td>
<td>Clinical trial network</td>
<td>222</td>
<td>222 (100)</td>
<td>0</td>
<td>0</td>
<td>All were members of clinical trial networks.</td>
<td>46 (North America, South America, Europe, Asia, Australasia, Africa)</td>
</tr>
<tr>
<td>Serrano</td>
<td>Identified by patient groups</td>
<td>53</td>
<td>0</td>
<td>53 (100)</td>
<td>0</td>
<td>All participants were patients</td>
<td>1 (Europe)</td>
</tr>
<tr>
<td>Taylor 2005</td>
<td>Clinical trial network</td>
<td>32</td>
<td>32 (100)</td>
<td>0</td>
<td>0</td>
<td>Most were members of a clinical trial network</td>
<td>10 (North America, Europe, Australasia, Africa)</td>
</tr>
<tr>
<td>Taylor 2008</td>
<td>Clinical trial network</td>
<td>33</td>
<td>30 (91)</td>
<td>0</td>
<td>0</td>
<td>Most participants were members of OMERACT</td>
<td>11 (North America, South America, Europe, Asia, Australasia)</td>
</tr>
<tr>
<td>Zochling</td>
<td>Clinical trial network</td>
<td>55</td>
<td>53 (96)</td>
<td>2 (4)</td>
<td>0</td>
<td>All were members of a clinical trial network</td>
<td>Unclear (unclear)</td>
</tr>
</tbody>
</table>

**Footnotes to Table 6:**

a: anaesthetists, social scientists and lactation specialists (numbers of each group unknown)

b: 23 clinicians completed all 3 rounds. Unclear how many completed first round
Anonymity

In thirteen studies conducted by email, post or internet, seven were conducted completely anonymously (participants were not aware who the other members of the group were, and individuals’ answers were not shared with the group) (Devane et al 2007; Distler et al 2008; Lux, & Osborne 2004; McGrath et al 2008; Ruperto et al 2003; Serrano-Aguilar et al 2009; Taylor et al 2008), and in the others complete anonymity is presumed (it is unclear whether participants knew the identities of other individuals). In the two studies in which the Delphi process was conducted at face-to-face meetings, voting was anonymous in one (Dent et al 2008) but not the other (Khan, & Pallant 2007).

In nine studies, participants met before (Miller et al 2001; Taylor et al 2008), during (Dent et al 2008; Khan, & Pallant 2007), after (Douglas et al 2009; McGrath et al 2008; Ruperto et al 2003; Zochling et al 2008), or before and after (Mease et al 2008) the Delphi process.

The level of anonymity in each study is summarised in Table 7.
Table 7 Level of anonymity in the studies, and whether participants met at any point

<table>
<thead>
<tr>
<th>Study</th>
<th>Number of rounds</th>
<th>How Delphi was conducted</th>
<th>Did participants meet</th>
<th>Did participants know the identity of other group members?</th>
<th>Did participants know the answers provided by other individuals in the group?</th>
<th>Level of anonymity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dent</td>
<td>3</td>
<td>Meeting</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Quasi</td>
</tr>
<tr>
<td>Devane</td>
<td>3</td>
<td>Internet</td>
<td>No</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Complete</td>
</tr>
<tr>
<td>Distler</td>
<td>3</td>
<td>Internet</td>
<td>No b</td>
<td>No</td>
<td>No</td>
<td>Complete</td>
</tr>
<tr>
<td>Douglas</td>
<td>3</td>
<td>Email</td>
<td>Yes c</td>
<td>Unclear</td>
<td>No</td>
<td>Complete</td>
</tr>
<tr>
<td>Khan</td>
<td>3</td>
<td>Meeting</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Not</td>
</tr>
<tr>
<td>Khanna</td>
<td>3</td>
<td>Email</td>
<td>No b</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Complete</td>
</tr>
<tr>
<td>Lux</td>
<td>6</td>
<td>Email</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Complete</td>
</tr>
<tr>
<td>Mease 2008</td>
<td>Unclear</td>
<td>Yes d</td>
<td>Unclear</td>
<td>No</td>
<td>Presumed</td>
<td>Complete</td>
</tr>
<tr>
<td>McGrath</td>
<td>2</td>
<td>Email</td>
<td>Yes c</td>
<td>No</td>
<td>No</td>
<td>Complete</td>
</tr>
<tr>
<td>Miller</td>
<td>2</td>
<td>Email</td>
<td>Yes c</td>
<td>Unclear</td>
<td>No</td>
<td>Presumed</td>
</tr>
<tr>
<td>Ruperto</td>
<td>2</td>
<td>Postal</td>
<td>Yes c</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Complete</td>
</tr>
<tr>
<td>Serrano</td>
<td>3</td>
<td>Email</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Complete</td>
</tr>
<tr>
<td>Taylor 2005</td>
<td>3</td>
<td>Email</td>
<td>No</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Complete</td>
</tr>
<tr>
<td>Taylor 2008</td>
<td>3</td>
<td>Internet</td>
<td>Yes e</td>
<td>No</td>
<td>No</td>
<td>Complete</td>
</tr>
<tr>
<td>Zochling</td>
<td>3</td>
<td>Email</td>
<td>Yes f</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
</tr>
</tbody>
</table>

Footnotes to Table 7:

a: at least part of the Delphi process was conducted at a meeting
b: steering group met but did not participate in the Delphi process
c: Results of Delphi process were used to inform participants at a subsequent consensus meeting
d: Delphi was preceded by focus groups which did not constitute part of the Delphi process itself
e: Delphi was preceded by meetings which did not constitute part of the Delphi process itself
f: The results of the Delphi process were presented at a conference, involving participants, and the list of core outcomes was refined
Structure of the Delphi process

The general format involved identification of potential outcomes, followed by determination of those that were felt to be most important or appropriate for clinical trials.

Identification of potential outcomes

In four studies, participants suggested outcomes, at the start of the Delphi process, without prompting or guidance from the facilitators (Douglas et al 2009; McGrath et al 2008; Ruperto et al 2003; Serrano-Aguilar et al 2009). In two studies, participants suggested outcomes within a framework of domains, suggested by a steering group (Khanna et al 2008), or based on international recommendations (Khan, & Pallant 2007).

In four studies, outcomes were proposed by a steering committee (Distler et al 2008), or by facilitators who, by reviewing the literature, identified outcomes used in previous studies (Lux, & Osborne 2004; Taylor et al 2008; Zochling et al 2008). In three studies, outcomes measured in clinical trials were discussed at international meetings, and a list of potentially eligible outcomes for the Delphi surveys were identified (Miller et al 2001; Dent et al 2008; Taylor et al 2008). In one study, outcomes were identified from both a systematic review of clinical trials and by asking participants to suggest two ‘new’ outcomes that were not listed (Devane et al 2007). In another, clinicians considered outcomes suggested by a steering group, and patients considered outcomes identified in focus groups.
Determining the importance of potential outcomes

Participants either scored the importance of each outcome (Devane et al. 2007; Distler et al. 2008; Douglas et al. 2009; Khanna et al. 2008; Serrano-Aguilar et al. 2009; Taylor et al. 2008), voted for or against its measurement in clinical trials (Miller et al. 2001; Dent et al. 2008; Khan, & Pallant 2007; Lux, & Osborne 2004; Zochling et al. 2008), distributed a set number of points amongst outcomes, according to importance (Mease et al. 2008; Taylor 2005), or ranked outcomes in order of importance (McGrath et al. 2008; Ruperto et al. 2003). In two studies, participants were asked to justify their answers (Khan, & Pallant 2007; Lux, & Osborne 2004).

Feedback of the results to participants after each round

Nine groups fed back either the average score for each outcome (Devane et al 2007; Distler et al 2008; Douglas et al 2009; Mease et al 2008; Serrano-Aguilar et al 2009; Taylor 2005; Taylor et al 2008), or the percentage of people voting for its inclusion in the core set (Khan, & Pallant 2007). In four studies, facilitators analysed data and presented a new list of outcomes, without presenting a measure of group opinion to participants Khanna 2008; Lux 2004; Mcgrath 2008; Ruperto 2003). In two
studies it was unclear what measure of group response was fed back to participants (Miller et al 2001; Zochling et al 2008).

**How consensus was reached about which outcomes to measure**

Eight studies used the Delphi technique as the main method of reaching consensus about which outcomes to measure, of which six recommended outcomes which received a pre-determined score (Khanna et al 2008; Taylor et al 2008), or a pre-determined proportion of participants felt it should be included in a core set (Dent et al 2008; Khan, & Pallant 2007; Lux, & Osborne 2004; Zochling et al 2008). These scores and proportions varied between the six studies. In another study (Devane et al 2007), an outcome was included in the core set if its score was higher than the mean score of all outcomes and at least 70% of participants scored it 4/5 on a Likert-type scale. In one study, a steering group determined the final core set of outcomes (Distler et al 2008).

Seven studies did not use the Delphi process itself to reach consensus about which outcomes to measure. Six of these used the results of the Delphi to inform people participating in subsequent consensus studies that aimed to design core outcome sets. Two of these studies (Douglas et al 2009; Ruperto et al 2003) used the Delphi process as a way of filtering out outcomes felt to be less important, and these were not considered at the subsequent consensus meeting. In the other four studies (Miller et al 2001; McGrath et al 2008; Mease et al 2008; Taylor 2005), each outcome
considered in the Delphi process was carried forward to the subsequent consensus meeting, regardless of its score. The aim of one study was to inform the design of a systematic review (Serrano-Aguilar et al 2009), and all the outcomes that were considered are ranked in order of importance, but the final decision about whether they should be measured in the review is not described.

The decisions about which of the outcomes that were initially suggested should be measured in clinical research studies are summarised in Table 8.
Table 8 How consensus was reached about which outcomes to measure in clinical trials

<table>
<thead>
<tr>
<th>Study</th>
<th>Was Delphi the final method for reaching consensus, or was it followed by another consensus process</th>
<th>Definition of consensus</th>
<th>Initial number of outcomes/domains</th>
<th>Number of outcomes/domains recommended</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dent</td>
<td>Delphi was used to write consensus document about RCT design</td>
<td>Consensus was reached if at least 75% of participants agreed with a statement that an outcome should be measured</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Devane</td>
<td>Delphi was final method for reaching consensus</td>
<td>After each round, only outcomes whose mean score was higher than the mean score of all outcomes, and at least 70% of participants scored it 4/5 on a Likert-type scale, were carried forward. After the final round, outcomes which met this criteria were included in the core set</td>
<td>263 outcomes</td>
<td>48 outcomes</td>
</tr>
<tr>
<td>Distler</td>
<td>Delphi was final method for reaching consensus</td>
<td>The core set was selected by a steering group, after the final round of the Delphi. This was done using cluster analysis, in which outcome domains suggested by the group as being most important were classed as feasible or not</td>
<td>17 domains</td>
<td>7 domains</td>
</tr>
<tr>
<td>Douglas</td>
<td>Delphi process was followed by a meeting at which NGT was used to determine the final core set</td>
<td>After the first and second rounds, outcomes with a median score of 3/9 or less on a Likert-type scale were removed. After the final round, only outcomes with a median score of at least 6/9 on a Likert-type scale were considered in a subsequent consensus meeting. At this meeting, 80% of participants had to agree for an outcome to be included in the core set.</td>
<td>220 outcomes</td>
<td>40 outcomes</td>
</tr>
<tr>
<td>Khan</td>
<td>Delphi was final method for reaching consensus</td>
<td>After Round 1, outcomes voted for by &lt; 50% of participants were removed. After the final round, outcomes voted for by &gt; 50% of participants included in core set</td>
<td>144 outcomes</td>
<td>30 outcomes</td>
</tr>
<tr>
<td>Was Delphi the final method for reaching consensus, or was it followed by another consensus process</td>
<td>Definition of consensus</td>
<td>Initial number of outcomes/domains</td>
<td>Number of outcomes/domains recommended</td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------------------------------------------------------</td>
<td>------------------------</td>
<td>-----------------------------------</td>
<td>---------------------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>Khanna</strong></td>
<td>Delphi was final method for reaching consensus</td>
<td>After the first two rounds, a 9 person steering committee filtered out outcomes they felt were not appropriate. After the second round of the Delphi process, this steering committee considered each outcome, and chose whether it should still be considered. Outcomes rejected by at least 33% of participants were removed. In Round 3, all participants considered each outcome and scored, on a scale of 1-9, whether it should be included in the core set (9=should be included). Outcomes with a median score of at least 7 were included in the core set.</td>
<td>212 outcomes</td>
<td>31 outcomes</td>
</tr>
<tr>
<td><strong>Lux</strong></td>
<td>Delphi was used to write consensus document about trial design (including statements about outcomes)</td>
<td>There was no pre-defined level of consensus about statements relating to which outcomes should be measured. Instead, the participants agreed on the overall recommendations at the end of the Delphi process.</td>
<td>n/a $^b$</td>
<td>n/a $^b$</td>
</tr>
<tr>
<td><strong>Mease</strong></td>
<td>Results of Delphi process were used to help participants at subsequent meetings decide which outcome domains to measure</td>
<td>There were no pre-defined cut-off scores for domains to be considered, at the consensus meeting involving clinicians and researchers, for inclusion in the core set. The participants at that meeting suggested that outcome domains which at least 50% of participants felt to be important should be considered ‘key’ domains for clinical trials. The top 15 domains considered in the separate patient Delphi process were listed as ‘Top 15’ domains from patients.</td>
<td>40 domains presented to clinicians. 104 domains presented to patients</td>
<td>6 domains for acute pain. 8 domains for chronic pain</td>
</tr>
<tr>
<td><strong>McGrath</strong></td>
<td>Results of Delphi process were used to help participants at subsequent meetings decide which outcomes to measure</td>
<td>There were no pre-defined cut-off scores for domains to be considered at the consensus meeting. At this meeting, participants agreed on the overall recommendations at the end of the Delphi process.</td>
<td>Unclear for acute pain. 6 domains for chronic pain</td>
<td>6 domains for acute pain. 8 domains for chronic pain</td>
</tr>
<tr>
<td>Study</td>
<td>Methodology</td>
<td>Definition of consensus</td>
<td>Initial number of outcomes/domains</td>
<td>Number of outcomes/domains recommended</td>
</tr>
<tr>
<td>-------------</td>
<td>------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------</td>
<td>-----------------------------------</td>
<td>----------------------------------------</td>
</tr>
<tr>
<td>Miller</td>
<td>Results of Delphi process were used to help participants at subsequent meetings decide which outcomes to measure</td>
<td>There were no pre-defined cut-off scores for outcomes to be considered, at the consensus meeting, for inclusion in the core set. At this meeting, if at least 70% of participants felt an outcome should be measured in clinical research studies, it was included in final core set.</td>
<td>unclear</td>
<td>5 domains, 7 outcomes</td>
</tr>
<tr>
<td>Ruperto</td>
<td>Delphi process was followed by a meeting at which NGT was used to determine the final core set</td>
<td>Only outcomes suggested by at least 10 participants in Round 1 were carried forward to Round 2. There were no pre-defined cut-off scores for outcomes to be considered, at the consensus meeting, for inclusion in the core set. At this meeting, if at least 70% of participants felt an outcome should be measured in clinical research studies, it was included in final core set.</td>
<td>Unclear how many outcomes initially suggested. 41 outcomes for SLE and 37 for JDM were considered in Round 2.</td>
<td>SLE: 8 domains, 11 outcomes. JDM: 11 domains, 15 outcomes</td>
</tr>
<tr>
<td>Serrano</td>
<td>Delphi was used to rank the importance of outcomes</td>
<td>Outcomes were ranked in order of importance after participants scored each from 1-10.</td>
<td>11 outcomes</td>
<td>11 outcomes</td>
</tr>
<tr>
<td>Taylor 2005</td>
<td>Aim of Delphi was not to identify a core set of outcomes, but to inform future consensus meetings</td>
<td>There was no pre-defined level of consensus about which domains should be measured. The authors did split the domains into ‘higher’ and ‘lower’ scoring</td>
<td>26 domains (spanning all situations)</td>
<td>6 domains for rehabilitation trials, 7 domains for trials of disease controlling drugs, 5 domains for trials of disease modifying drugs</td>
</tr>
<tr>
<td>Taylor 2008</td>
<td>Delphi was final method for reaching consensus</td>
<td>Outcomes were scored out of 7 as follows: 1-3=should definitely include in core set, 4=uncertain, 7-9=should definitely not include. In Round 2, all outcomes from Round 1 were presented again. In Round 3, only outcomes for which there was disagreement (ie bimodal response) or a median score of 4. After Round 3, outcomes with a median score of 1-3, and for which there was no disagreement, were included in the core set.</td>
<td>7 domains for acute gout, 15 domains for chronic gout</td>
<td>6 domains for acute gout, 10 domains for chronic gout</td>
</tr>
</tbody>
</table>
Was Delphi the final method for reaching consensus, or was it followed by another consensus process

**Definition of consensus**

In the first two rounds, outcomes were removed if less than 80% of participants voted for its inclusion in the core set. In the final round, outcomes voted by at least 50% of participants were included in the core set

**Initial number of outcomes/domains**

7 domains

**Number of outcomes/domains recommended**

7 domains

<table>
<thead>
<tr>
<th><strong>Zochling</strong></th>
<th>Delphi was final method for reaching consensus</th>
<th>7 domains</th>
</tr>
</thead>
</table>

**Abbreviations used in Table 8:** JDM= Juvenile dermatomyositis; NGT= Nominal Group Technique; SLE=systemic lupus erythematosis

**Footnotes to Table 8:**

a: The aim of the study was to provide a list of statements about specific outcomes in trials of certain therapies. For example, “the primary outcome measure of a reflux chest pain syndrome trial should be a clinically meaningful reduction of chest pain”.

b: The aim of the study was not to develop a core outcome set, but mainly to identify the most appropriate primary outcome, and to provide some guidance about the measurement and reporting of important secondary outcomes

c: The primary aim of this study was to develop a list of core domains (called ‘elements’) that should be included in an international registry of patients with Ankylosing Spondylitis. 2/7 domains relate to demographic data and biologic-specific data, rather than outcomes of treatment. The remaining 5 domains relate to outcomes of treatment (clinical parameters, physical function, disease activity, imaging and Quality of Life)
Attrition of participants

In four studies, no participants who completed the first round subsequently dropped out of the study (Dent et al 2008; Khan, & Pallant 2007; McGrath et al 2008; Serrano-Aguilar et al 2009), and in one study the only participants who dropped out completed all but the final round of the study (Taylor 2005). In the remaining ten studies, some participants dropped out of the Delphi process before the final round. In five of these, (Distler et al 2008; Lux, & Osborne 2004; Mease et al 2008; Ruperto et al 2003; Zochling et al 2008) each participant was invited to every round, even if they did not complete previous questionnaires. In three studies (Devane et al 2007; Khanna et al 2008; Taylor et al 2008) participants who did not respond to any given round were excluded from the remainder of the study. In two studies, additional participants were invited as the Delphi progressed (Miller et al 2001; Douglas et al 2009).

4.5 Discussion

Across studies using the Delphi technique to identify which outcomes to measure in clinical research studies, the methodology differs considerably, and the quality of reporting of key aspects is variable. Researchers dominate participation in such studies, and patients and families are seldom involved.
Composition of the groups

Informed clinical decisions can only be based on the results of trials that have measured outcomes of importance to both clinicians and patients. Initiatives to identify which outcomes to measure in clinical trials, however, focus on the opinions of researchers. This means that outcomes included in existing core sets may be selected to serve the needs of researchers in academia or industry, rather than according to how important they are to patients.

Outcomes important to clinicians or researchers may differ from those used by patients. In one study, in which patients were involved in the design of a systematic review, they highlighted certain outcomes as particularly important, but these had not been measured in any of the included trials (Serrano-Aguilar et al 2009). Research conducted within the OMERACT group also suggests that clinicians and researchers may not realise that certain outcomes are very important for patients (Mease et al 2008). The perspective of patients is now routinely incorporated into the work conducted by OMERACT (Kirwan et al 2009).

The opinions of different groups can be analysed either together or separately. The use of multiple panels, each comprising a different group (Mease et al 2008), acknowledges that there may be differences in opinion. If different groups with potentially conflicting views are included in a single panel, they may not be equally represented in the final consensus. This can happen either because the panel includes more participants from a certain group, so the final consensus is numerically dominated by their
responses (Devane et al 2007), or because participants tailor their answers to agree with a group which they perceive to be more authoritative (Serrano-Aguilar et al 2009).

In studies that use a single panel, comprising a mixture of participants, authors should report a measure of the distribution of scores for each outcome considered in the final round. This is because cut-off scores, used in most studies, do not describe how strongly the minority feel, and so an apparent consensus could actually be masking major disagreement within the group (Goodman 1987).

**Avoiding bias introduced by researchers and facilitators**

So that researchers do not impose their views on participants, and thus introduce bias into the study, participants are traditionally asked open questions in the first round of a Delphi process. In the context of identifying which outcomes to measure in clinical research studies, this would translate into the practice of participants suggesting potential outcomes that they feel should be considered in the Delphi process, and would not be prompted or guided by facilitators, steering committees, or reviews of the literature. Most studies we identified did not take this approach. It is possible that provision of a list of outcomes to participants for initial consideration may overstate the importance of outcomes that are favourable to the researchers, rather than those which may be of more importance to clinicians and patients. It is known that outcomes measured
in previous clinical trials do not always reflect those deemed most appropriate by all stakeholders (Gandhi et al 2008; Duncan et al 2000).

**Avoiding bias occurring because of participant attrition**

People with minority opinions may be more likely to drop out of studies that use the Delphi process, so attrition as rounds progress can lead to overestimation of the degree of consensus in the final results. Strategies to prevent attrition bias are to only invite people who respond to a pre-Delphi invitation to participate in the first round (Ruperto et al 2003) or to list, in the final publication, only those participants who either completed the entire Delphi process or agreed the final consensus statement (Lux, & Osborne 2004).

**Suggested aspects of the methodology and results that should be reported**

When the Delphi technique is used for identifying which outcomes to measure in clinical research, authors should describe important methodological aspects of the process in the study report. This will help enable appraisal of the study, which may affect whether or not the recommendations are implemented.

Criticisms of the Delphi technique are that ‘expertise’ of the panel is arbitrarily defined, and the validity of the final consensus is questionable,
because individual participants are not accountable for their responses, and they may be led towards conformity with the group, rather than consensus of true opinions (Sackman 1975). It has also been suggested that people who drop out of studies that use the Delphi technique are more likely to hold opinions which differ from the majority view, and so the degree of consensus reached in the final round may be overestimated (Bardecki 1984).

To allow the reader to assess whether the composition of the panel was appropriate, authors should report the number and types of participants involved in the study, and how they were identified.

To enable assessment of the Delphi process itself, they should report how potential outcomes were identified at the start of the study, and the process of determining which of these should be measured in clinical trials. Such details must include the types of questions asked, and how authors decided whether consensus had been reached amongst the group members.

To enable assessment of the risk of bias from attrition of participants, authors should report whether or not non-responders were invited to continue in the study or whether they were excluded. They should also report the number of people who dropped out in each round.

As a minimum, the results should describe the group opinion for each outcome that was included in the final round. They should describe a measure of the group response (eg average score) and distribution (eg interquartile range).
Finally, given the variations between studies, it would be helpful if authors explained the rationale behind their methodological choices, or discussed the effects these may have on the results.

Robustness of this review

Half of the studies we identified related to rheumatological conditions. Similarities in the methods used in these studies may have skewed our findings towards those proposed by the OMERACT group. However, we did identify differences between the studies, in terms of methodological decisions and reporting quality.

Future areas of methodological research

Given variations in methodology between studies, we feel there is a need for research to determine how best to develop core outcome sets. An agenda for this research could be designed through the COMET initiative (Core Outcome Measures for Effectiveness Trials), which is an international network of individuals and organisations with interest or experience of the development, application and promotion of core outcome sets (http://www.liv.ac.uk/nwhtmr/comet/comet.htm). One such area of ongoing research and discussion relates to whether core outcome sets designed for clinical practice, such as those developed in the five studies we excluded (Brunner et al 2008; Jones et al 2000; van Hulst et al 2009;
Weigl et al 2004; Radtke et al 2009), should be the same as those designed for research. Another priority is research to identify the most effective ways to incorporate the views of different groups of participants, especially patients, in the design of core outcome sets.

4.6 Summary

Studies which use the Delphi process for the purpose of gaining consensus around a core outcome set for clinical trials are conducted to improve the quality and usefulness of clinical research, and in order for their recommendations to be implemented, they should be of sufficiently high quality. Where possible patients and clinicians should be involved in developing core outcome sets, and they should be directly asked, initially using open questioning techniques, which outcomes are of particular importance in routine clinical practice.

Poor reporting of methodological decisions makes it difficult to appraise these studies. Researchers wishing to use the Delphi technique for this purpose should be aware of these issues when designing their study, and explain methodological decisions, in relation to the study aims, in the main publication.

The need for a core outcome set or childhood asthma was identified and discussed in Chapter 3. One important part of the process of developing such a core set is to identify which outcomes are most important and
relevant. A study, in which the Delphi process was utilised to identify these outcomes, is described in the next chapter. We used this systematic review, in which we identified methodological factors that should strengthen the validity of the Delphi technique, and others that put the process at risk of bias, to help design the study described in Chapter 5.
Chapter 5

Identification of the most relevant outcomes for evaluating regular therapies for childhood asthma

5.1 Background

In order to inform clinical practice, late phase clinical trials should measure outcomes that are important and relevant to patients, clinicians, and policymakers. As we identified in Chapter 2, very little work has been conducted to identify which outcomes are important and relevant in childhood asthma. In one study, 14 clinicians and researchers indicated, by questionnaire, which outcomes they felt were most appropriate for a variety of clinical, public health and research settings related to asthma (Smith et al 1996).

As shown in the systematic review in Chapter 3, outcomes in RCTs in children with asthma are focussed on short-term disease activity, but other domains are much less frequently measured. Furthermore, there is inconsistency between studies in the way that outcomes are measured and reported. For example, exacerbations are defined in several different ways, and a variety of scales are used to assess symptoms.
We recommended that one solution to these problems would be to agree a minimum set of core outcomes that should be measured, and reported, in all late phase RCTs in children with asthma. This core outcome set should be aligned to the assessments that are used, in clinical consultations, to decide whether a treatment regime is satisfactory, or whether it should be modified, and should include a few outcomes that are particularly important to patients, families, and clinicians.

5.2 Aim

To identify which outcomes are most important when evaluating regular therapies for childhood asthma, from the perspective of the following people: clinicians involved in the out-patient management of children with asthma; parents of children younger than 18 years, who have asthma; and young people with asthma, who are aged between 13 and 18 years.

5.3 Methodological approach

Identifying the outcomes used when making shared decisions in clinic consultations

In previous attempts to develop core outcomes sets, participants have, generally, been asked to consider outcomes that were already measured in clinical trials. One problem that may arise from this approach is that some
important outcomes, if they are not routinely measured in clinical trials, may be overlooked. Conversely, this approach may perpetuate the inclusion in core sets of other outcomes, which may not be so relevant, on the basis that they have traditionally been measured in clinical trials.

Rather than take this approach, we elected not to start this study by presenting to the participants the frequency with which outcomes were measured in clinical trials. Instead, we focussed on identifying which outcomes are important for making shared decisions in clinic consultations. During these consultations, clinicians, parents, and some young people themselves, develop an overall assessment of whether the current therapy regime is satisfactory, or whether it should be modified, and we felt that these outcomes should be reflected in clinical trials. To our knowledge, this approach, of using the Delphi process to identify outcomes, for clinical trials, by asking participants to focus only on how they assess treatments in clinical practice, has not been utilised before.

We also acknowledged that the opinions of young people, parents and clinicians may differ, and so we felt that their results should be ascertained and analysed separately. This avoids bias occurring when groups with conflicting views are analysed together, if one group contains more people, or certain groups of participants tailor their answers to agree with the views of others whom they perceive to be more authoritative.
School-age and pre-school children

We considered outcomes for pre-school (less than 5 years) and school-aged (5 to 18 years) children separately, because we anticipated that differences between the two age groups may lead to variation in the relevance of certain outcomes. Asthma in these two age groups differs phenotypically (Bush 2009), and in terms of the natural history (Kovesi et al 2010), and this may lead to different goals of therapies. Other important age-related differences may relate to the measurement of outcomes. Physiological tests of lung function are rarely measured in pre-school children, because of difficulties they may have with performing the procedure. Functional status may also be measured in different ways, because when children start attending school, the pattern of their normal daily activities changes.

5.4 Participants and methods

5.4.1 Structure of the study

This study was conducted, using questionnaires, in two phases (1 and 2). The purpose of Phase 1 was to identify a long list of potential outcomes, and Phase 2 was designed to identify which of these was most important. Participants were asked to consider outcomes for evaluating therapies taken regularly, rather than for acute exacerbations.

To ascertain the views of clinicians, a two-round, web-based, anonymised Delphi survey was conducted amongst members of the British Paediatric
Respiratory Society (BPRS), and a network of asthma nurses. The BPRS comprises medical and non-medical professionals, within the UK, who care for children with respiratory problems. To be eligible for the study, clinicians and nurses needed to have ongoing experience of managing children with asthma, but they did not require experience of designing or conducting clinical trials.

Parents were invited to complete paper-based surveys in asthma clinics in Alder Hey Children’s Hospital (AHCH), a large paediatric hospital in the North of England, which takes referrals from primary and secondary care. Parents of all children younger than 18 years, who were prescribed regular preventer therapy for asthma, and did not have respiratory co-morbidities, were eligible. Young people aged 13 to 18, who attended these clinics, were also invited, because we anticipated that they might have different goals for their asthma treatment. The lower age limit was based on clinical experience that teenagers are generally more able to discuss their asthma than younger children.

5.4.2 Phase 1

Phase 1 comprised the first round of the Delphi survey of clinicians, and a survey of parents and young people. Open questions were asked, in order to identify a long list of outcomes that could be relevant in clinic consultations.
Phase 1 clinician questionnaire

The questionnaire distributed to clinicians in Phase 1, including the initial invitation email sent through the BPRS, is included in Appendix 6. To avoid introducing researcher bias, by prompting responses based on our own ideas, we asked open questions to identify which outcomes clinicians use in clinic consultations. We asked the following question once in relation to pre-school children, and once for school-aged children:

“When you see children with asthma in clinic, you make an assessment as to whether their treatment is working. Please list up to five beneficial or harmful outcomes of treatment that you find clinically most important in school aged/pre-school children. These factors should be things that you consider, when deciding whether to recommend continuing on current treatment or altering a child’s regular asthma therapy regime.”

Phase 1 parent and young peoples’ questionnaires

The questions on the initial drafts of the parents’ and young peoples’ questionnaires were initially devised after discussion amongst IS, PRW and RLS. General comments on the layout and wording of the first draft of the questionnaire for young people were sought from the MCRN Young Peoples’ group, who regularly assess clinical trial information leaflets designed for his age group.
Four versions of the parent and young people’s questionnaires were piloted before the study. This was done in asthma clinics, where parents and young people were asked their opinions about the wording of the questions. The main comment about terminology was that parents felt that the phrase “are you happy with your child’s asthma control” referred to their satisfaction with health care professionals and health service provision, so this was amended.

The Phase 1 questionnaire for parents is included in Appendix 7. Parents of pre-school and school-aged children were asked four exploratory questions. The first two questions aimed to identify which outcomes were used to assess whether a treatment regime was satisfactory or not:

1) “Over the last twelve months, have you generally felt that the regular preventer treatment that your child takes has kept their asthma under control? Yes/No. If you ticked YES, please tell us what aspects of your child’s asthma, or their daily life, have made you feel happy that they are on the correct regular medication. If you ticked NO, please leave this question blank.”

2) “Over the last twelve months, have there been times when you felt that your child’s regular preventer treatment should be increased or changed, because their asthma was not under control? Yes/No. If you ticked YES, please tell us the reasons why you were not satisfied with the regular preventer treatment that they were taking? If you ticked NO, please leave this space blank”.

The other questions aimed to identify other outcomes of potential importance, such as long term beneficial and harmful effects of treatment:

3) “Does anything worry you about the fact that your child has asthma? Yes/No. If you ticked YES, please tell us the worries you have about the fact your child has asthma. If you ticked NO, please leave this space blank”

4) “Does anything worry you about the regular preventer treatment that your child takes for their asthma? Yes/No. If you ticked YES, please tell us what worries you have about the treatment your child takes for their asthma. Please be as specific as you can. If you ticked NO, please leave this question blank.”

**Analysis of the Phase 1 questionnaires**

IS interpreted each response from clinicians, parents and young people, and decided which outcome of treatment was being described. The broad framework for classifying responses as outcomes was based on the domains, subdomains and outcomes identified in Chapter 3. This comprised the following six categories:

Short-term disease activity: Symptoms; relief inhaler use; exacerbations; lung function; overall asthma control

Physical consequences of disease: death; progression of asthma into later childhood or adulthood
Functional status: ability to exercise or play sport; activities of daily living; school attendance

QoL/family: overall QoL; emotional well-being; family outcomes

Adverse effects of therapy: short term adverse effects; long term adverse effects

Health resource utilisation

The reviewers (IS, RG, RLS, PRW) discussed whether each of the responses that did not fit into this classification should constitute a ‘new’ outcome.

To ensure that the views of each group were taken into account, we pre-specified that outcomes suggested by at least 10% of children and/or parents and/or clinicians would be carried forward and listed on the Phase 2 questionnaire. It was felt that a lower cut-off lower would have meant that too many outcomes would be carried forward, placing extra burden on participants. Outcomes were also considered for inclusion in the Phase 2 questionnaire, after discussion amongst reviewers, if they were measured in at least 10% of RCTs identified in Chapter 3, even if they had not been suggested by at least 10% of participants.

In order to identify responses that were open to interpretation, and to make categorisation of responses more accurate, Reviewer 2 (RG) independently analysed the first 36 (72%) questionnaires received from parents and young people, and Reviewer 3 (PRW) independently analysed a randomly selected
sample of 9/46 (20%) of the questionnaires completed by clinicians. Further review was not deemed necessary because interpretation of responses became easier as the study progressed. All disagreements were discussed, initially, with Reviewer 3 (PRW), and, if agreement was not reached, Reviewer 4 (RLS). In addition, RLS was shown all responses felt to be atypical or difficult to interpret.

**Phase 1 participants**

The study flowchart is shown in Figure 5.
Figure 5 Flowchart showing participants in Phases 1 and 2

**Clinician Delphi**
- 261 invited to Phase 1
- 46 completed Phase 1
- 1 dropped out

**Phase 1:**
Open questions to identify potential outcomes

Outcomes listed by 10% of participants carried forward

**Phase 2:**
Closed questions to identify most important outcomes

- 45 invited to Phase 2
- 43 completed Phase 2

**Parent surveys (2 cohorts)**
- 52 invited to Phase 1
- 49 completed Phase 1
- 3 declined

**Parent surveys (2 cohorts)**
- 51 invited to Phase 2
- 1 declined

Incl. young people:
- 27 parents of school-age children
  - 17/27 aged 5-12
  - 10/27 aged 12-18
  - 11/27 aged 13-18
- 11 young people
  - all aged 13-18

**Parent surveys (2 cohorts)**
- 21 parents of pre-school children
  - 18/21 aged 3-5
  - 3/21 aged 5-7

**Parent surveys (2 cohorts)**
- 2/23 aged under 5
  - 13/23 aged 1-5
The email invitation was sent to 260 members of the BPRS, and 21 specialist asthma nurses. Of these 281 invitees, 46 participated in Stage 1 of the study. This group comprised 38 hospital-base paediatricians and 8 nurses. Of the doctors, 14/38 stated that they were consultant respiratory paediatricians, 13/38 were consultant general paediatricians, 3/38 were consultant paediatricians with a specialist interest in respiratory paediatrics, 6/38 were clinical academics (4 professors and 2 senior lecturers), 1/38 was a clinical research fellow and 1/38 was a specialist registrar in a national training post in respiratory paediatrics. One nurse was a specialist health visitor, and the others were based in hospitals. The institutions at which the participants worked encompassed a mixture of district general hospitals and tertiary paediatric centres, distributed throughout the United Kingdom.

Of 44 asthmatic children identified in clinics, three were not eligible. In all three cases, this was because the child had concomitant respiratory diagnoses (two had immunodeficiencies predisposing to recurrent lower respiratory tract infection and one had bronchiolitis obliterans). In total, 52 people (41 parents and 11 young people) were approached to participate in the study. Of the parents, three declined (one mother did not have time, and one grandmother and one father did not live with the children they had brought to clinic, so did not feel as if they could complete the questionnaire). All 11 young people who were approached agreed to participate, including the young person whose grandmother declined.

Of the participants, 27/49 (55%) were parents of school-aged children (of whom 17 were aged at least five years but not older than 12 years, and 10
were older than 12 years but younger than eighteen years), 11/49 (22%) were parents of pre-school children (all of whom were older than one year of age, but younger than five years), and 11/49 (22%) were young people (ranging from 13 to 15 years). The ages of the children ranged from one year to fifteen years.

5.4.3 Phase 2

Only the clinicians who participated in Phase 1 were invited to complete the Phase 2 questionnaire. Parents attending asthma clinics at AHCH were invited to participate, employing the same eligibility criteria as those used in Phase 1. To increase numbers, parents of pre-school children were recruited at a local district general hospital, whose paediatric respiratory department treats children with asthma with broadly similar characteristics to those seen at AHCH. As it was unclear whether young people would be able to answer the questionnaire in Phase 2, they were not included. In another attempt to increase the size of the sample, the Phase 2 questionnaire was distributed by Asthma UK, the largest UK-based asthma charity, to parents of children with asthma, who had previously expressed an interest to be involved in clinical research.
Phase 2 questionnaires

The questionnaire used in Phase 2 were designed by IS, RLS and PRW, and shown to clinicians within the department for their comments. The clinician and parent questionnaires are shown in Appendix 8.

In the clinician questionnaire, outcomes carried forward for school-aged and pre-school children were listed separately, and the questions were repeated for both age groups. Parents received a questionnaire in which outcomes for both age groups were listed, and they were instructed only to score those relevant to their own child.

Participants were asked the following three questions:

1) “Regular treatments for children can have a variety of beneficial effects, each of which could be measured as an outcome in clinical trials. Please score how important each of the following outcomes are on a scale of 0-4.” This question aimed to identify the relative importance of each outcome.

2) “You have just scored the outcomes listed below. Please tick the three most important”.

3) “Are there any other outcomes which are not listed that would have made it into your top three?”
Analysis of the Phase 2 questionnaires

The scores from clinicians and parents were analysed separately. Scores from parents of pre-school children were analysed separately from those of school-aged children. For each outcome, we calculated the following:

Median score and IQR; the proportion of participants selecting the outcome in the top 3

The median score was used as an overall measure of whether the outcome was important to the group. The main parameter for ranking the outcomes, by age-group, was the proportion of participants who selected the outcome in their top 3.

Participants in Phase 2

Of the 46 clinicians who participated in Phase 1, 45 agreed to participate in Phase 2. 43/45 (96%) participants who were sent the link to the Phase 2 questionnaire responded. One of the non-responders was a clinical research fellow, and the other was a consultant respiratory paediatrician. The answers they provided in Phase 1 were similar to those of the overall group.

Of a possible 53 parents whose child’s case notes were reviewed, two were not eligible. In both cases, this was because the child had concomitant respiratory diagnoses (one had immunodeficiency predisposing to lower respiratory tract infection, and one had bronchiolitis obliterans). Of the 51 parents who were invited to participate, nobody declined. One parent,
however, took the questionnaire away but did not return it. In total, therefore, 50 parent questionnaires were analysed.

Of parents who completed the Phase 2 questionnaire, 27/50 had attended clinic with a school-aged child (of whom 18 were aged at least 5 years but not older than 12 years, and 9 were aged at least 12 years but not older than 18 years), and 23/50 attended with a pre-school child (of whom 2 were younger than one year and 21 were at least one year but not older than five years). Eight of these were recruited at APH. The children ranged from less than one to 16 years of age.

In spite of several reminders, we received only 13/118 (11%) responses from Asthma UK. For this reason, these data have not been included in our analyses. They are, however, summarised in Appendix 9.

5.4.4 Ethical considerations

Details of the study were emailed to the National Research Ethics Service (NRES), who categorised it as “service evaluation and development”, which did not require ethical review by an NHS Research Ethics Committee. A copy of this correspondence is included in Appendix 10. The main ethical considerations are described below.

The study was conducted according to the General Medical Council’s guidance about consent (GMC 2008). Young people, parents and clinicians
were clearly told that, although it would be helpful if they participated in the study, they were not obliged to.

**Data protection for clinicians**

Clinician questionnaires were handled according to the eight principles described in the Data Protection Act (1998) (http://www.opsi.gov.uk/acts/acts1998/ukpga_19980029_en_9#sch1-pt1 – accessed 20th June 2010).

Clinicians were told that their participation would be acknowledged in the final study report. Their email addresses were held securely and not distributed to any third party. Of several online survey websites evaluated before Phase 2, Survey Gizmo appeared to be one of the most established and reputable (http://www.surveygizmo.com/ - accessed 16th November 2009). A password is required to retrieve the information provided by participants. Furthermore, the administrators do not ‘harvest’ email addresses, or access data provided by participants.

**Confidentiality for patients and parents**

We did not ask for any details by which patients could be identified. Completed questionnaires were placed in an opaque box in the out-patient department. In order to avoid discussing the child’s diagnosis of asthma in the waiting room, patients were approached discreetly.
5.5 Results

5.5.1 Outcomes suggested in Phase 1

Clinicians suggested 21 separate outcomes for pre-school children, and 21 for school-aged children. Parents of pre-school children suggested 15 outcomes, parents of school-aged children suggested 19, and young people suggested 13. Outcomes for pre-school and school-aged children were broadly similar.

In total, participants suggested 29 different outcomes, of which symptoms, effects of asthma on the ability to exercise and function as the child would like, and adverse effects of therapy, were the most frequently suggested. The frequency with which outcomes were suggested, and the domains into which they were categorised, are shown for clinicians considering pre-school children and school-aged children, parents of pre-school children, parents of school-aged children, and young people, in Tables 9, 10, 11, 12 and 13 respectively.
Table 9 Outcomes suggested in Phase 1 by clinicians, for pre-school children

<table>
<thead>
<tr>
<th>Domain</th>
<th>Outcome</th>
<th>Number (%) of clinicians suggesting the outcome (n=46)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Disease Activity</strong></td>
<td>Daytime symptoms</td>
<td>43 (93)</td>
</tr>
<tr>
<td></td>
<td>Nocturnal symptoms</td>
<td>28 (61)</td>
</tr>
<tr>
<td></td>
<td>Symptoms affecting physical activity</td>
<td>17 (37)</td>
</tr>
<tr>
<td></td>
<td>Need for bronchodilator</td>
<td>23 (50)</td>
</tr>
<tr>
<td></td>
<td>Exacerbations</td>
<td>29 (63)</td>
</tr>
<tr>
<td></td>
<td>Patient/parent overall view</td>
<td>4 (8)</td>
</tr>
<tr>
<td></td>
<td>Ability to feed</td>
<td>2 (4)</td>
</tr>
<tr>
<td></td>
<td>Wheezing in response to allergens</td>
<td>3 (7)</td>
</tr>
<tr>
<td></td>
<td>Normal chest examination</td>
<td>3 (7)</td>
</tr>
<tr>
<td><strong>Disease Damage</strong></td>
<td>Death</td>
<td>1 (2)</td>
</tr>
<tr>
<td><strong>Functional status</strong></td>
<td>Activities of daily living</td>
<td>9 (20)</td>
</tr>
<tr>
<td></td>
<td>School/nursery</td>
<td>5 (11)</td>
</tr>
<tr>
<td><strong>Quality of Life/family</strong></td>
<td>Overall Quality of Life</td>
<td>4 (9)</td>
</tr>
<tr>
<td></td>
<td>Family outcomes</td>
<td>8 (17)</td>
</tr>
<tr>
<td><strong>Adverse effects of therapy</strong></td>
<td>Short term adverse effects</td>
<td>10 (22)</td>
</tr>
<tr>
<td></td>
<td>Growth</td>
<td>8 (17)</td>
</tr>
<tr>
<td></td>
<td>Other long term adverse effects</td>
<td>1 (2)</td>
</tr>
<tr>
<td><strong>Health resource utilisation</strong></td>
<td>Unscheduled visit</td>
<td>9 (18)</td>
</tr>
<tr>
<td></td>
<td>Hospital admission</td>
<td>20 (45)</td>
</tr>
<tr>
<td></td>
<td>Admission to intensive care unit</td>
<td>1 (2)</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td>Developmental milestones</td>
<td>1 (2)</td>
</tr>
</tbody>
</table>
Table 10 Outcomes suggested in Phase 1 by clinicians for school-aged children

<table>
<thead>
<tr>
<th>Domain</th>
<th>Outcome</th>
<th>Number (% of clinicians suggesting the outcome (n=46)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease Activity</td>
<td>Daytime symptoms</td>
<td>43 (92)</td>
</tr>
<tr>
<td></td>
<td>Nocturnal symptoms</td>
<td>30 (63)</td>
</tr>
<tr>
<td></td>
<td>Symptoms affecting physical activity</td>
<td>29 (63)</td>
</tr>
<tr>
<td></td>
<td>Need for bronchodilator</td>
<td>27 (58)</td>
</tr>
<tr>
<td></td>
<td>Exacerbations</td>
<td>31 (67)</td>
</tr>
<tr>
<td></td>
<td>Patient/parent overall view</td>
<td>4 (8)</td>
</tr>
<tr>
<td></td>
<td>Lung function</td>
<td>14 (30)</td>
</tr>
<tr>
<td></td>
<td>Inflammatory markers (^a)</td>
<td>2 (4)</td>
</tr>
<tr>
<td></td>
<td>Normal chest exam</td>
<td>2 (4)</td>
</tr>
<tr>
<td></td>
<td>Change in FEV1 in response to Salbutamol</td>
<td>2 (4)</td>
</tr>
<tr>
<td>Disease Damage</td>
<td>--No outcomes suggested--</td>
<td></td>
</tr>
<tr>
<td>Functional status</td>
<td>Activities of daily living</td>
<td>14 (30)</td>
</tr>
<tr>
<td></td>
<td>School attendance</td>
<td>27 (59)</td>
</tr>
<tr>
<td>Quality of Life/family</td>
<td>Overall Quality of Life</td>
<td>7 (15)</td>
</tr>
<tr>
<td></td>
<td>Family outcomes</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Adverse effects of therapy</td>
<td>Short term adverse effects</td>
<td>6 (13)</td>
</tr>
<tr>
<td></td>
<td>Growth</td>
<td>11 (24)</td>
</tr>
<tr>
<td></td>
<td>Other long term adverse effects</td>
<td>15 (30)</td>
</tr>
<tr>
<td>Health resource utilisation</td>
<td>Unscheduled visit</td>
<td>7 (15)</td>
</tr>
<tr>
<td></td>
<td>Hospital admission</td>
<td>13 (28)</td>
</tr>
<tr>
<td>Other</td>
<td>Achieving parent/ patient-defined goal</td>
<td>1 (2)</td>
</tr>
</tbody>
</table>

Footnotes:
\(^a\): both clinicians suggested Exhaled Nitric Oxide (ENO)
Table 11 Frequency with which outcomes were suggested by parents of pre-school children (n=11)

<table>
<thead>
<tr>
<th>Domain</th>
<th>Outcome</th>
<th>Number (%) of parents of preschool children suggesting outcome (n=11)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Disease activity</strong></td>
<td>Daytime symptoms</td>
<td>8 (73)</td>
</tr>
<tr>
<td></td>
<td>Nocturnal symptoms</td>
<td>1 (9)</td>
</tr>
<tr>
<td></td>
<td>Symptoms affecting activity (eg exercise)</td>
<td>3 (27)</td>
</tr>
<tr>
<td></td>
<td>Need for reliever medication (eg bronchodilator)</td>
<td>2 (18)</td>
</tr>
<tr>
<td></td>
<td>Exacerbations</td>
<td>4 (36)</td>
</tr>
<tr>
<td></td>
<td>Patient/parent overall view on asthma control</td>
<td>1 (9)</td>
</tr>
<tr>
<td></td>
<td>Ability to feed</td>
<td>1 (9)</td>
</tr>
<tr>
<td><strong>Disease damage</strong></td>
<td>--No responses categorised into this domain--</td>
<td></td>
</tr>
<tr>
<td><strong>Functional status</strong></td>
<td>ADL</td>
<td>2 (18)</td>
</tr>
<tr>
<td></td>
<td><strong>DOMAIN:</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Quality of life/family</strong></td>
<td>Overall QoL</td>
<td>1 (9)</td>
</tr>
<tr>
<td></td>
<td>Emotional well being</td>
<td>1 (9)</td>
</tr>
<tr>
<td><strong>Adverse effects of therapy</strong></td>
<td>Short term AE</td>
<td>5 (45)</td>
</tr>
<tr>
<td></td>
<td>Growth</td>
<td>1 (9)</td>
</tr>
<tr>
<td></td>
<td>Other long term AE</td>
<td>4 (36)</td>
</tr>
<tr>
<td></td>
<td>Dependence</td>
<td>1 (9)</td>
</tr>
<tr>
<td><strong>Health resource utilisation</strong></td>
<td>--No responses categorised into this domain--</td>
<td></td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td>Health-related problems when child is older</td>
<td>2 (18)</td>
</tr>
<tr>
<td>Domain</td>
<td>Outcome</td>
<td>Number (%) of parents of school-age children suggesting outcome (n=27)</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>-------------------------------------------------</td>
<td>---------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Disease activity</strong></td>
<td>Daytime symptoms</td>
<td>9 (33)</td>
</tr>
<tr>
<td></td>
<td>Nocturnal symptoms</td>
<td>20 (74)</td>
</tr>
<tr>
<td></td>
<td>Symptoms affecting physical activity</td>
<td>11 (41)</td>
</tr>
<tr>
<td></td>
<td>Need for reliever medication (eg bronchodilator)</td>
<td>3 (11)</td>
</tr>
<tr>
<td></td>
<td>Exacerbations</td>
<td>13 (48)</td>
</tr>
<tr>
<td></td>
<td>Patient/parent overall view</td>
<td>5 (19)</td>
</tr>
<tr>
<td></td>
<td>Ability to feed</td>
<td>1 (4)</td>
</tr>
<tr>
<td><strong>Disease damage</strong></td>
<td>Death</td>
<td>3 (11)</td>
</tr>
<tr>
<td><strong>Functional status</strong></td>
<td>Activities of daily living</td>
<td>13 (48)</td>
</tr>
<tr>
<td></td>
<td>School/nursery attendance</td>
<td>2 (7)</td>
</tr>
<tr>
<td><strong>Quality of life/family</strong></td>
<td>Overall Quality of Life</td>
<td>4 (15)</td>
</tr>
<tr>
<td></td>
<td>Emotional well being</td>
<td>1 (4)</td>
</tr>
<tr>
<td><strong>Adverse effects of therapy</strong></td>
<td>Short term adverse effects</td>
<td>16 (60)</td>
</tr>
<tr>
<td></td>
<td>Growth</td>
<td>2 (7)</td>
</tr>
<tr>
<td></td>
<td>Other long term adverse effects</td>
<td>14 (52)</td>
</tr>
<tr>
<td></td>
<td>Development of dependence on medication</td>
<td>2 (7)</td>
</tr>
<tr>
<td><strong>Health resource utilisation</strong></td>
<td>A+E/GP visit</td>
<td>2 (7)</td>
</tr>
<tr>
<td></td>
<td>Hospital admission</td>
<td>2 (7)</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td>Health-related problems when child is older</td>
<td>5 (19)</td>
</tr>
</tbody>
</table>
Table 13 Frequency with which outcomes were suggested by young people (n=11)

<table>
<thead>
<tr>
<th>Domain</th>
<th>Outcome</th>
<th>Number (%) of parents of young people suggesting outcome (n=11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease activity</td>
<td>Daytime symptoms</td>
<td>5 (45)</td>
</tr>
<tr>
<td></td>
<td>Nocturnal symptoms</td>
<td>2 (18)</td>
</tr>
<tr>
<td></td>
<td>Symptoms affecting physical activity</td>
<td>4 (36)</td>
</tr>
<tr>
<td></td>
<td>Exacerbations</td>
<td>1 (9)</td>
</tr>
<tr>
<td>Disease damage</td>
<td>Death</td>
<td>1 (9)</td>
</tr>
<tr>
<td>Functional status</td>
<td>Activities of daily living</td>
<td>3 (27)</td>
</tr>
<tr>
<td>Quality of life/family</td>
<td>Overall quality of life</td>
<td>1 (9)</td>
</tr>
<tr>
<td></td>
<td>Emotional well being</td>
<td>1 (9)</td>
</tr>
<tr>
<td>Adverse effects of therapy</td>
<td>Short term adverse effects</td>
<td>2 (18)</td>
</tr>
<tr>
<td></td>
<td>Growth</td>
<td>1 (9)</td>
</tr>
<tr>
<td></td>
<td>Other long term adverse effects</td>
<td>3 (27)</td>
</tr>
<tr>
<td>Health resource utilisation</td>
<td>Hospital admission</td>
<td>1 (9)</td>
</tr>
<tr>
<td>Other</td>
<td>Health-related problems when child is older</td>
<td>1 (9)</td>
</tr>
</tbody>
</table>
5.5.2 Agreement between reviewers assessing the Phase 1 questionnaires

Interpretation of answers from parents and young people: independent analysis by IS and reviewer 2 (RG)

Reviewer 2 was shown a sample of 36 completed questionnaires, from 27 parents and 9 young people. Reviewer 1 felt that, from the responses in these 36 questionnaires, participants had suggested a total of 124 items that could be interpreted as representing an outcome, and Reviewer 2 felt they had suggested 102. Of the 124 items that Reviewer 1 had extracted, Reviewer 2 agreed with 100 (81%). In total, there were 26 disagreements, of which 24/26 occurred when Reviewer 1 had categorised a response as an outcome but Reviewer 2 had not, and 2/26 had occurred when Reviewer 2 had categorised a response as an outcome but reviewer 1 had not.

Interpretation of answers from clinicians: independent analysis by IS and reviewer 3 (PRW)

Reviewer 3, who was not aware of IS’s interpretation of the individual questionnaires, was shown a randomly selected sample of 9/46 clinician questionnaires (20%), and found complete agreement with this evaluation.
Arbitration of the disagreements between IS and Reviewer 2: Discussion with Reviewers 3 (PRW) and 4(RLS)

Reviewer 3 assessed all 26 of these disagreements. With regard to 20/26 (77%), she agreed with IS. On the 6/26 occasions (23%) that she agreed with Reviewer 2, IS had classified a response as an outcome but Reviewer 2 did not. IS discussed each of these six responses with Reviewer 3.

After discussion about one response, IS agreed with Reviewer 3. This related to a parent’s description that their child “saw [the] asthma nurse who confirmed that his results through breathing tests are much better”. IS initially thought this meant that the parents felt lung function was an important outcome. After discussion with Reviewer 3, IS agreed that the parents were simply reporting what the nurse did, rather than saying lung function is an important outcome.

The other five disagreements were arbitrated by Reviewer 4, who agreed with IS’s assessment about four of these. IS had thought the outcome “overall asthma control” was most appropriate for the following responses from parents: “[she] has her asthma well controlled through the medication she takes”; “we do not feel it controls [his] asthma”; “her asthma is more under control”; “well controlled”. Reviewer 3 had suggested that the comments were not detailed enough to interpret in this way. Reviewer 4 suggested IS’s classification was appropriate, because “overall asthma control” is a concept that parents often discuss in out-patient consultations.
One disagreement related to whether outcomes relating to exercise should be categorised in the functional status domain (ie ability to exercise) or the disease activity domain (ie exercise-related symptoms). After discussion with Reviewer 4, IS agreed that this outcome is more appropriately placed in the disease activity domain.

**Discussion of atypical comments between IS and Reviewer 4 (RLS)**

Reviewer 4 was asked to check 18 responses, which IS felt may be open to interpretation. IS and Reviewer 2 had agreed on the classification of these responses. Reviewer 4 deemed the classification to be correct on 16/18 occasions. After discussion, IS agreed with Reviewer 4 that the classification should be changed for the other two responses.

For one of these, Reviewer 2 and IS had interpreted the response “the way my child can be normal and then fighting for breath” as relating to an exacerbation of asthma, but Reviewer 4 suggested this was just a description of acute shortness of breath, a frequent interval symptom which does not, in itself, constitute an exacerbation. For the other, Reviewer 2 and IS had classified “staying on medication for a long period” as representing a worry about having asthma later in life. Reviewer 4 disagreed with this classification, and suggested that this is more likely to represent a concern about long-term complications of asthma therapy.
Interpretation of responses which did not fit exactly into the initial outcomes framework: discussion amongst IS, Reviewer 3 and Reviewer 4

Symptoms

Symptoms were described in a variety of ways. Some parents responses described cough, wheeze, and shortness of breath. These were all classified as “symptoms”, rather than listing them separately on the Phase 2 questionnaire. Where parents specified that these were nocturnal problems, they were classed as such. Therefore, symptoms were classed as either nocturnal or daytime. Some parents discussed symptoms as part of a response to upper respiratory tract infections. It was felt that these could represent exacerbations of asthma, and were classed as such.

Risk of problems in the future

Responses relating to long-term respiratory or overall health, and future impact of asthma on children’s lives, were felt to be related. It was difficult to determine whether participants were referring to future risk of asthma, other respiratory illness, general health problems, or problems with functional status. It was also difficult to discern whether ‘future risk’ referred to later childhood or adulthood, so we felt it would be inappropriate to divide these into separate outcomes based on our interpretation of responses. These were, therefore, combined into the outcome ‘health related problems when older’. This outcome was not classed into any of the pre-defined domains, as it was felt to span several of them, but rather in a separate ‘other’ category.
**Adverse effects of medication**

Adverse effects of medications were grouped together into three outcomes, namely short-term problems, growth, and other long-term problems. Short-term problems included non-systemic adverse effects, and other comments about side-effects, unless they were specified by participants as relating to growth or other long-term problems. Long-term problems included systemic side effects, and responses from participants that described a future risk from medications because of long-term use.

**Quality of life**

Certain responses from parents were interpreted as referring to overall quality of life, because they either alluded to parents describing a child’s overall well-being, or discussed a combination of functional and emotional problems from having asthma. IS and Reviewer 2 had complete agreement that these should be categorised as quality of life. Reviewer 4 had seen all these responses, and also agreed that they described overall quality of life. These responses are shown below:

(1) “[She] is able to live her day-to-day life (school and home) as a normal child”;  

(2) “I worry that as she gets older her asthma gets worse and will prevent her from leading a normal life”;  

(3) “His general day-to-day quality of life”;
(4) “Difficulties in controlling symptoms, problematic breathing having impact on aspects of daily living, sports, socialising with friends and sleep. I am concerned my son has had anxiety due to worrying about asthma”

(5) “I feel happy because I can do more”;

(6) “Not as tired during the day and not as bad tempered. Eating habits have greatly improved – generally more engaged in life”

Responses that were not classed as outcomes

Some suggestions were not categorised as outcomes of treatment, because they did not relate to whether a treatment works, or does more good than harm.

Four clinicians and one set of parents suggested that the ease which parents can administer the asthma therapy, and whether they have the correct technique for using inhalers, are important outcomes. We felt that, although these may affect the decision about whether to change a treatment modality, they do not reflect the efficacy or safety of therapy.

Five clinicians suggested compliance with treatment to be an important outcome. We felt this not to be an outcome of treatment, as a variety of factors determine whether parents and young people comply with a therapy regime, not just whether it is felt to be efficacious and safe.
One young person said she was worried about whether, in the future, her children will be at risk of having asthma. We did not consider this to be an outcome associated with asthma treatments.

One clinician suggested that, in clinic appointments with pre-school children, it is important to use licensed medications.

**Final list of outcomes included in the questionnaire in Phase 2**

The final list of outcomes carried forward to Phase 2, and those which were not, is shown in Table 14.
Table 14 Whether outcomes were carried forward to Phase 2: determined by the frequencies with which outcomes were suggested by participants at the end of Phase 1, and with which they were measured in RCTs of inhaled corticosteroids for children with asthma.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Clinicians School age children (n=46)</th>
<th>Clinicians Pre school children (n=46)</th>
<th>Parents of school age children (n=27)</th>
<th>Young people (n=11)</th>
<th>Parents of preschool children (n=11)</th>
<th>RCTs including school age children (n=95)</th>
<th>RCTs including preschool children (n=47)</th>
<th>Is outcome carried forward for school age children</th>
<th>Is outcome carried forward for pre-school age children</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DISEASE ACTIVITY</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Daytime symptoms</td>
<td>43 (92)</td>
<td>43 (93)</td>
<td>9 (33)</td>
<td>5 (45)</td>
<td>8 (73)</td>
<td>84 (88)</td>
<td>43 (91)</td>
<td>Yes</td>
<td>Yes</td>
<td>84/95 and 43/47 RCTs measured symptoms. Daytime, nocturnal and exercise-induced symptoms were included in these scores. Because of the frequency with which this outcome was measured in RCTs in pre-school children, it was carried forward to the Phase 2 questionnaire for pre-school parents</td>
</tr>
<tr>
<td>Nocturnal symptoms</td>
<td>30 (63)</td>
<td>28 (61)</td>
<td>20 (74)</td>
<td>2 (18)</td>
<td>1 (9)</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Symptoms affecting activity (eg exercise)</td>
<td>29 (63)</td>
<td>17 (37)</td>
<td>11 (41)</td>
<td>4 (36)</td>
<td>3 (27)</td>
<td></td>
<td></td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Need for bronchodilator</td>
<td>27 (58)</td>
<td>23 (50)</td>
<td>3 (11)</td>
<td>0</td>
<td>2 (18)</td>
<td>71 (75)</td>
<td>39 (83)</td>
<td>Yes</td>
<td>Yes</td>
<td>This outcome includes suggestions by clinicians, that exacerbations, specifically in response to URTIs, are an important outcome</td>
</tr>
<tr>
<td>Exacerbations</td>
<td>31 (67)</td>
<td>29 (63)</td>
<td>13 (48)</td>
<td>1 (9)</td>
<td>4 (36)</td>
<td>31 (33)</td>
<td>17 (36)</td>
<td>Yes</td>
<td>Yes</td>
<td>Although parents did not specifically describe this outcome, 5 parents did describe terms relating to overall asthma control</td>
</tr>
<tr>
<td>Patient/parent overall view on asthma control</td>
<td>3 (6)</td>
<td>3 (6)</td>
<td>5 (19)</td>
<td>0</td>
<td>1 (9)</td>
<td>9 (9)</td>
<td>6 (13)</td>
<td>Yes</td>
<td>Yes</td>
<td>Not carried forward for preschool children because it was suggested by no parents or clinicians</td>
</tr>
<tr>
<td>Lung function</td>
<td>14 (30)</td>
<td>0</td>
<td>1 (4)</td>
<td>0</td>
<td>0</td>
<td>89 (94)</td>
<td>36 (77)</td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Bronchoconstriction induced by challenge</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>28 (30)</td>
<td>1 (2)</td>
<td>No</td>
<td>No</td>
<td>Not carried forward for school-age children because it was suggested by no parents or clinicians</td>
</tr>
<tr>
<td>Inflammatory markers</td>
<td>2 (4)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>16 (17)</td>
<td>3 (6)</td>
<td>No</td>
<td>No</td>
<td>Not carried forward (said by 2/46 clinicians; no parents)</td>
</tr>
<tr>
<td>Ability to feed</td>
<td>0</td>
<td>2 (4)</td>
<td>1 (4)</td>
<td>0</td>
<td>1 (9)</td>
<td>*</td>
<td>*</td>
<td>No</td>
<td>No</td>
<td>*May have been measured as part of symptom score in RCTs, but not as an outcome in itself</td>
</tr>
<tr>
<td>Occurrence of wheeze in response to allergens</td>
<td>0</td>
<td>3 (7)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>*</td>
<td>*</td>
<td>No</td>
<td>No</td>
<td>*May have been measured as part of symptom score in RCTs, but not as an outcome in itself</td>
</tr>
<tr>
<td>Normal chest exam</td>
<td>2 (4)</td>
<td>3 (7)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Change in FEV1 in response to B2+</td>
<td>2 (4)</td>
<td>3 (7)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Outcome</td>
<td>Clinicians School age children (n=46)</td>
<td>Clinicians Pre school children (n=46)</td>
<td>Parents of school age children (n=27)</td>
<td>Young people (n=11)</td>
<td>Parents of preschool children (n=11)</td>
<td>RCTs including school age children</td>
<td>RCTs including pre-school children</td>
<td>Included for school age children</td>
<td>Included for preschool age children</td>
<td>Notes</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
<td>----------------------------------------</td>
<td>----------------------------------------</td>
<td>---------------------------------------</td>
<td>---------------------</td>
<td>-------------------------------------</td>
<td>----------------------------------</td>
<td>-----------------------------------</td>
<td>----------------------------------</td>
<td>-----------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td><strong>DISEASE DAMAGE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Death</td>
<td>0</td>
<td>1 (2)</td>
<td>3 (11)</td>
<td>1 (9)</td>
<td>0</td>
<td>*</td>
<td>*</td>
<td>Yes</td>
<td>Yes</td>
<td>*Measured implicitly but not reported</td>
</tr>
<tr>
<td>Lung growth</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1 (1)</td>
<td>0</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Development milestones</td>
<td>1 (2)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td><strong>FUNCTIONAL STATUS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ADL</td>
<td>14 (30)</td>
<td>9 (20)</td>
<td>13 (48)</td>
<td>3 (27)</td>
<td>2 (18)</td>
<td>6 (6)</td>
<td>5 (10)</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>School/nursery</td>
<td>27 (59)</td>
<td>5 (11)</td>
<td>2 (7)</td>
<td>0</td>
<td>0</td>
<td>13 (14)</td>
<td>4 (4)</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>QoL/FAMILY OUTCOMES</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall QoL</td>
<td>7 (15)</td>
<td>4 (9)</td>
<td>4 (15)</td>
<td>1 (9)</td>
<td>1 (9)</td>
<td>7 (7)</td>
<td>5 (11)</td>
<td>Yes</td>
<td>Yes</td>
<td>Incorporator in QoL scores, but not measured as an outcome itself</td>
</tr>
<tr>
<td>Emotional well being</td>
<td>0</td>
<td>0</td>
<td>1 (4)</td>
<td>1 (9)</td>
<td>1 (9)</td>
<td>*</td>
<td>*</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Family outcomes</td>
<td>1 (2)</td>
<td>8 (17)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7 (7)</td>
<td>10 (21)</td>
<td>No</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td><strong>ADVERSE EFFECTS OF THERAPY</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short term AE</td>
<td>6 (13)</td>
<td>10 (22)</td>
<td>16 (60)</td>
<td>2 (18)</td>
<td>5 (45)</td>
<td>78 (82)*</td>
<td>43 (96)*</td>
<td>Yes</td>
<td>Yes</td>
<td>* Did not distinguish between short and long term AEs</td>
</tr>
<tr>
<td>Growth</td>
<td>11 (24)</td>
<td>8 (17)</td>
<td>2 (7)</td>
<td>1 (9)</td>
<td>1 (9)</td>
<td>30 (32)</td>
<td>16 (34)</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Other long term AE</td>
<td>15 (30)</td>
<td>1 (2)</td>
<td>14 (52)</td>
<td>3 (27)</td>
<td>4 (36)</td>
<td>*</td>
<td>*</td>
<td>Yes</td>
<td>Yes</td>
<td>* Did not distinguish between short and long term AEs</td>
</tr>
<tr>
<td>Dependence</td>
<td>0</td>
<td>0</td>
<td>2 (7)</td>
<td>0</td>
<td>1 (9)</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td><strong>HEALTH RESOURCE UTILISATION</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A+E/GP visit</td>
<td>7 (15)</td>
<td>9 (18)</td>
<td>2 (7)</td>
<td>0</td>
<td>0</td>
<td>8 (8)</td>
<td>7 (15)</td>
<td>Yes</td>
<td>Yes</td>
<td>* Either included in exacerbations or unscheduled HRU</td>
</tr>
<tr>
<td>Hospital admission</td>
<td>13 (28)</td>
<td>20 (45)</td>
<td>2 (7)</td>
<td>1 (9)</td>
<td>0</td>
<td>*</td>
<td>*</td>
<td>Yes</td>
<td>Yes</td>
<td>* Either included in exacerbations or unscheduled HRU</td>
</tr>
<tr>
<td><strong>OTHER</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health-related problems when child is older</td>
<td>0</td>
<td>0</td>
<td>7 (26)</td>
<td>1 (9)</td>
<td>2 (18)</td>
<td>0</td>
<td>0</td>
<td>Yes</td>
<td>Yes</td>
<td>Includes disease damage (ie progression of respiratory illness into older childhood or adulthood) and long term functional status (eg career, lifestyle when older)</td>
</tr>
<tr>
<td>Achieving parent/ patient-defined goal</td>
<td>1 (2)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>No</td>
<td>No</td>
<td></td>
</tr>
</tbody>
</table>

*: Issues relating to the reporting of outcomes in RCTs are described in the column entitled ‘notes’
Abbreviations: ADL=Activities of Daily Living; AE=Adverse effect; A+E=Accident and Emergency; B2+= Beta agonist; FEV1= Forced Expiratory volume in 1 second; ICS= Inhaled corticosteroids; PEFR=Peak Expiratory Flow Rate; QoL=Quality of life; URTI=upper respiratory tract infection
5.5.3 Phase 2: Ranking of outcomes in Phase 2

Almost all outcomes received a median score of 4 from either parents or clinicians. Generally, clinicians and parents agreed on which outcomes were important. Both groups rated symptoms, exacerbations, QoL, and mortality as the most important outcomes, for both age groups. Parents rated health-related problems when their child is older as a more important outcome than clinicians did.

The results for each outcome listed in the Phase 2 questionnaire, for preschool children, are shown in Table 15, and for school-aged children, are shown in Table 16.
Table 15 Preschool children: the importance of each outcome listed in the Phase 2 questionnaire, as scored by clinicians and parents.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Clinicians (n=43)</th>
<th>Parents (n=23)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Clinician Rank</td>
<td>Number (%)</td>
</tr>
<tr>
<td></td>
<td>Median (IQR)</td>
<td>scoring outcome</td>
</tr>
<tr>
<td></td>
<td></td>
<td>in Top 3</td>
</tr>
<tr>
<td>Nocturnal symptoms</td>
<td>1</td>
<td>4 (3,4)</td>
</tr>
<tr>
<td>Exacerbations</td>
<td>2</td>
<td>4 (3,4)</td>
</tr>
<tr>
<td>Quality of Life</td>
<td>3</td>
<td>3 (3,4)</td>
</tr>
<tr>
<td>Daytime symptoms</td>
<td>4</td>
<td>4 (3,4)</td>
</tr>
<tr>
<td>Death</td>
<td>5</td>
<td>4 (2,4)</td>
</tr>
<tr>
<td>Hospital admission</td>
<td>6</td>
<td>4 (3,4)</td>
</tr>
<tr>
<td>Parent/child global assessment of</td>
<td>7</td>
<td>4 (3,4)</td>
</tr>
<tr>
<td>control</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Impact of asthma on the family</td>
<td>8</td>
<td>4 (3,4)</td>
</tr>
<tr>
<td>Use of reliever</td>
<td>9</td>
<td>3 (3,4)</td>
</tr>
<tr>
<td>Normal activities</td>
<td>10</td>
<td>4 (3,4)</td>
</tr>
<tr>
<td>Long term AE</td>
<td>11</td>
<td>4 (3,4)</td>
</tr>
<tr>
<td>School attendance</td>
<td>12</td>
<td>3 (3,4)</td>
</tr>
<tr>
<td>Activity or exercise</td>
<td>13</td>
<td>3 (3,4)</td>
</tr>
<tr>
<td>GPA/E attendance</td>
<td>14</td>
<td>4 (3,4)</td>
</tr>
<tr>
<td>Growth</td>
<td>15</td>
<td>3 (3,4)</td>
</tr>
<tr>
<td>Health related problems when</td>
<td>16</td>
<td>3 (2,4)</td>
</tr>
<tr>
<td>older</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short term AE</td>
<td>17</td>
<td>3 (2,5)</td>
</tr>
</tbody>
</table>

Footnote: Young people were not involved in Phase 2
Table 16 School-age children: The importance of each outcome listed in the Phase 2 questionnaire, as scored by clinicians and parents.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Clinicians (n=43)</th>
<th>Parents (n=27)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rank (n=17)</td>
<td>Median (IQR)</td>
</tr>
<tr>
<td>Quality of Life</td>
<td>1</td>
<td>4 (3,4)</td>
</tr>
<tr>
<td>Exacerbations</td>
<td>2</td>
<td>4 (4,4)</td>
</tr>
<tr>
<td>Nocturnal symptoms</td>
<td>3</td>
<td>4 (4,4)</td>
</tr>
<tr>
<td>Normal activities</td>
<td>4</td>
<td>4 (4,4)</td>
</tr>
<tr>
<td>Daytime symptoms</td>
<td>5</td>
<td>4 (4,4)</td>
</tr>
<tr>
<td>Death</td>
<td>6</td>
<td>4 (2,4)</td>
</tr>
<tr>
<td>Use of reliever inhaler</td>
<td>7</td>
<td>4 (3,4)</td>
</tr>
<tr>
<td>Hospital admission</td>
<td>8</td>
<td>4 (3,4)</td>
</tr>
<tr>
<td>Activity or exercise</td>
<td>9</td>
<td>4 (3,4)</td>
</tr>
<tr>
<td>Long term adverse effect</td>
<td>10</td>
<td>4 (4,4)</td>
</tr>
<tr>
<td>School attendance</td>
<td>11</td>
<td>4 (3,4)</td>
</tr>
<tr>
<td>Parent/child global assessment of control</td>
<td>12</td>
<td>4 (3,4)</td>
</tr>
<tr>
<td>Lung function tests</td>
<td>13</td>
<td>3 (2,3)</td>
</tr>
<tr>
<td>GP/A+E attendance</td>
<td>14</td>
<td>4 (3,4)</td>
</tr>
<tr>
<td>Growth</td>
<td>15</td>
<td>4 (3,4)</td>
</tr>
<tr>
<td>Health related problems when older</td>
<td>16</td>
<td>3 (2,4)</td>
</tr>
<tr>
<td>Short term adverse effect</td>
<td>17</td>
<td>3 (2,3)</td>
</tr>
</tbody>
</table>

Footnote: Young people were not involved in Phase 2
5.6 Discussion

5.6.1 Main findings

We found that the most important outcomes in childhood asthma are symptoms, exacerbations, quality of life, and mortality. We identified these by asking clinicians, parents, and young people, how they evaluate benefits and harms of therapies when making shared decisions in clinic consultations. This differs from the usual approach, when developing core outcome sets, of focusing on outcomes measured in clinical trials.

The frequency and severity of symptoms and exacerbations are universally accepted factors for deciding whether a patient’s asthma is well controlled (Kroegel, & Wirtz 2009), and it is not surprising that we found them to be very important to parents and clinicians, when evaluating beneficial effects of interventions. These clinical measures of disease activity appear more important, when making clinical decisions, than physiological measures of lung function. Although lung function is frequently assessed in clinical trials, because it is an objective evaluation of efficacy, we found that parents and clinicians are guided far more by clinical measures of asthma control, which reflect effectiveness of therapy.

The effects of interventions on the ability to perform normal activities, play sport and not miss school were considered very important outcomes, as was overall quality of life. Unless an intervention improves these aspects of a child’s life, it is likely to be regarded, by parents and clinicians, as unhelpful. Studies which have examined the correlation between clinical
markers of asthma severity and quality of life have shown that an overall assessment of well-being “cannot be imputed from clinical outcomes, and that it must be measured and interpreted independently” (Juniper et al 2004), because clinical markers of disease do not help evaluate psychosocial and functional effects of having asthma, which are important determinants of a child’s overall QoL (Reichenberg, & Broberg 2000; Mattsson 1972; Collins et al 2008; van den Bemt et al 2010). Outcomes reflecting QoL are rarely assessed in clinical trials, but we would recommend that they should be measured, either individually, or as part of a validated composite outcome.

We found that parents place more emphasis on long-term beneficial effects of therapy than clinicians do. Previous qualitative research has identified that many parents of children with asthma worry about long-term effects on health (Schulz et al 1994), and long-term cohort studies show that asthma often persists into later childhood and adulthood (Morgan et al 2005; Oswald et al 1994). Long-term harmful effects of therapies were important to clinicians and parents in our study. These have also previously been shown to be a major concern for parents and children (Townsend et al 1991), and have been identified as a research priority in asthma (James Lind Alliance, 2007). We feel that, despite difficulties of measuring long-term effects of therapies, they should be regarded as important outcomes in clinical trials in children with asthma. Further work should be conducted to identify the most appropriate ways to measure and report these outcomes.
Another complication of asthma, which we found to be of particular importance for parents, was mortality. Around half of the parents who participated in Phase 2 selected ‘prevention of death’ as one of their top 3 outcomes, compared with a quarter of clinicians. One reason for this discrepancy may be that some clinicians think that because mortality from childhood asthma is rare, it is not particularly relevant when considering beneficial effects of therapies. We recommend, however, that death should be measured, as a secondary outcome, in both treatment and control groups in clinical trials in children with asthma.

5.6.2 Comparison with the ATS/ERS outcomes taskforce

A recent collaboration, led by the ATS and ERS, suggested standardized definitions of asthma control, severity and exacerbations, and recommended outcomes that could be used to measure these in clinical trials in asthma (Reddel et al 2009). The study was mainly focused on outcomes for trials in adolescents and adults, but the authors suggest that their outcomes are relevant for children older than six years of age, with some special considerations.

The authors conducted literature reviews, to identify outcomes measured in clinical trials in asthma. Seven working groups, composed of clinicians, researchers, personnel from drug regulatory authorities, and representatives from the pharmaceutical industry, discussed the suitability of these outcomes for evaluating, in clinical trials, current asthma control and future
asthma-related problems. Three working groups focused on the measurement and reporting of exacerbations, physiological outcomes and indirect measures of asthma control (i.e., unscheduled use of health care, loss of work and school productivity, and need for additional medication), and four discussed the use of composite scores, biomarkers, asthma control diaries and quality-of-life questionnaires. Two paediatricians were involved, with the task of assessing whether the group’s recommendations were applicable to clinical trials in children. Consensus was reached after round-table open discussions.

The outcomes suggested by the group were considered in two broad domains, namely “current control” and “future risk”. The recommended outcomes are listed in Table 17. In Table 18, the top six outcomes from our study are compared with the recommendations of the ATS/ERS group.
Table 17 Outcomes recommended by the ATS/ERS, categorised by whether they should be considered essential, desirable or optional

<table>
<thead>
<tr>
<th></th>
<th>Essential</th>
<th>Desirable</th>
<th>Optional</th>
</tr>
</thead>
</table>
| **For assessing current control** | 1. Symptom-free days  
2. Reliever use  
3. Composite scores  
4. Exacerbation (within last month)  
5. Quality of life | 1. FEV1  
2. Symptom/ reliever/ lung function diary  
3. Corticosteroid use  
4. Health care utilization |                                                                                          |
|                      |                                                                           | 1. FEV1  
2. Symptom/reliever/ lung function diary  
3. Corticosteroid use  
4. Health care utilization |                                                                                          |
| **For assessing future risk** | 1. Exacerbations  
2. Post-bronchodilator FEV1x (for assessment of lung function decline)  
3. Composite scores which assess functional status and symptoms (eg Asthma Control Questionnaire)  
4. Treatment side-effects  
5. Pre-bronchodilator FEV1 (as predictor for exacerbations) | 1. Symptom/reliever/ lung function diary  
2. Health care utilization  
3. Mortality due to asthma  
4. Airway hyperresponsiveness  
5. Biomarkers (sputum eosinophil, exhaled Nitric Oxide) | 1. Biomarkers  
2. Airway hyperresponsiveness |

Adapted from Reddel et al 2009
Table 18 Comparison between the top 6 outcomes for school-aged children, identified after Phase 2 of our study, and the recommendations of the ATS/ERS taskforce

<table>
<thead>
<tr>
<th></th>
<th>Our study: Top 6 outcomes suggested by parents/clinicians</th>
<th>Reddel et al (essential)</th>
<th>(Reddel et al optional)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symptoms</td>
<td>√</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Exacerbations</td>
<td>√</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>QoL</td>
<td>√</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Death</td>
<td>√</td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>Normal activities</td>
<td>√</td>
<td>As part of composite asthma score</td>
<td></td>
</tr>
<tr>
<td>Exercise ability</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reliever use</td>
<td></td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Lung function</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment side effects</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Healthcare utilisation</td>
<td></td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>Biomarkers</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bronchial hyperresponsiveness</td>
<td></td>
<td></td>
<td>√</td>
</tr>
</tbody>
</table>
As can be seen in Table 18, the recommendations of the ATS/ERS taskforce are comparable with our results, especially with regard to symptoms, exacerbations, effects of asthma on daily life, and adverse effects of therapy. We would recommend that because these outcomes were noted to be of particular importance in both studies, they should strongly be considered for inclusion in a core outcome set for childhood asthma. Methodological research, to identify the most appropriate tools to measure these important outcomes, should be conducted. This should be used to inform consensus about how, and when, these outcomes should be measured and reported in all clinical trials of regular therapies for children with asthma.

One difference between the studies was that physiological measures of lung function, though ‘essential’ outcomes in the ATS/ERS core set, were found to be one of the least important outcomes in our study. This may be because our study focussed on clinical trials in children, in whom the problems with measuring and interpreting lung function have already been discussed, whereas the ATS/ERS study primarily focussed on outcomes for trials in adults. Similar problems, relating to whether or not outcomes measured in clinical trials in adults are applicable to children, have been highlighted by other groups involved in designing core outcome sets (Griffiths et al 2005). The discrepancy between our recommendations and those of the ATS may also reflect our empirical approach of considering the importance of outcomes in the context of a clinic consultation, compared to the other group’s approach of considering outcomes primarily from the perspective of clinical trials.
5.6.3 Robustness of the study

Strengths of the study

Our methodological approach was based on the idea that outcomes measured in clinical research studies should reflect those that are important and relevant in clinic consultations. In particular, two methodological choices enabled us to take this approach. Firstly, we involved parents, young people, and paediatricians, regardless of whether they had been involved in the design or conduct of clinical trials. Secondly, in Phase 1, we elected to ask participants open questions to identify they outcomes they felt were clinically important, rather than presenting them with a list of outcomes that are already measured in clinical trials. We are confident that these choices enabled us to identify outcomes of particular relevance when making shared decisions in routine practice, even if they are not routinely measured in clinical trials. In addition, this approach helped us avoid participants rating outcomes as important simply because they are traditionally measured in clinical trials. We would recommend that this approach is used when developing core outcome sets in other conditions.

When designing this study, we were aware of sources of bias, and we addressed these accordingly. If participants had scored outcomes on the basis of how frequently they were measured in previous clinical trials, certain outcomes, namely death and long-term problems, would not have been included in the Phase 2 questionnaire. We avoided this bias by asking open questions, without prompting responses, in Phase 1. By revealing neither the identities of participants, nor their answers, to the group,
individuals did not tailor their answers to agree with those of others, which was especially important given the mix of experience and professional backgrounds in the group. By keeping the clinician questionnaires as brief as possible, and by sending reminders to complete the Phase 2 questionnaire, we observed minimal attrition of participants. This is especially important when using the Delphi technique because if participants with opinions that do not match those of the rest of the group drop out, the overall results can be a biased representation of the original group’s opinion.

Much of the validity of the study rested on the results of Phase 1, which determined the outcomes listed in Phase 2. We felt these results were valid and reliable for four reasons. Firstly, parents and young people offered extensive input when designing the questionnaires, and so we were confident that our questions, and the aim of the study, would be clearly understood. Secondly, the agreement between the four reviewers was good, which suggests that our interpretation of participants’ responses was correct. Thirdly, all the outcomes we carried forward to Phase 2 received a median score of at least 3, and most scored 4, from both parents and clinicians, which suggests that the outcomes we identified in Phase 1 were all of some importance. Fourthly, we felt that the list of outcomes presented to participants in Phase 2 was comprehensive, because participants, when asked, did not suggest other outcomes that were not listed. This was particularly reassuring, given that we elected to collect data using questionnaires, rather than interviews or focus groups. Although these other qualitative research techniques can generate ‘rich’ data, they are
more time and labour intensive, for researchers and participants, and require specialist training.

Another strength of our study was that we ensured that the views of each group, regardless of the number of participant, were considered equally. We carried outcomes forward from Phase 1 to Phase 2 if they had been suggested by 10% of either clinicians, parents, or young people, and so each group had equal opportunity to influence the composition of the Phase 2 questionnaire. In Phase 2, we analysed the responses of parents and clinicians separately, rather than as one homogenous panel of participants, so that we could identify outcomes of importance to both groups, and also so that we could note relevant discrepancies.

Weaknesses of the study

One weakness of the study is that the sample is focussed on children whose asthma is managed by hospital paediatricians, although much asthma is treated in primary care and the community. Phase 2 of the study could be replicated amongst GPs, practice nurses, and parents with children whose asthma is managed in primary care. We feel, however, that our results are generalisable with regard to the types of children included in clinical trials, which are frequently conducted in the hospital setting.

Despite good agreement between reviewers, certain responses from parents in Phase 1 were open to interpretation. One example related to the description of symptoms. Although some parents specified whether they
were referring to daytime or nocturnal symptoms, others did not. A pragmatic decision was taken that a description of cough, wheeze, or shortness of breath should be classified as referring to nocturnal symptoms only if the response was this specific. There were two potential errors that could have occurred, if our assumption had been false. Firstly the number of participants suggesting ‘daytime symptoms’ could have been overestimated, and this outcome could have erroneously been included in the Phase 2 questionnaire. Secondly the number of participants suggesting ‘night-time symptoms’ could have been underestimated, and it could have been erroneously missed from the Phase 2 questionnaire. Ultimately, our results appear to be robust to this decision, because nocturnal symptoms were included in the Phase 2 questionnaire, and daytime symptoms scored highly.

Responses from parents completing the Phase 1 questionnaire, which were classified as representing quality of life, were also open to interpretation. Although there are a variety of different schemes for describing quality of life, and its constituent components, we decided to keep a broad definition that reflects a patient’s overall sense of well being, including psychosocial status. Even if we had misinterpreted these responses in Phase 1, the composition of the Phase 2 questionnaire would not have been affected, as the term ‘quality of life’ was suggested by a sufficient number of clinicians for school-aged children, and was measured in a sufficient proportion of clinical trials of pre-school children, to warrant its inclusion for both age groups. A detailed examination of what participants meant by quality of life was outside the scope of this study.
A final weakness is that the response rate of clinicians initially was poor at the start of the study. This was in contrast with our experiences with parents and young people, almost all of whom agreed to participate. If this study were to be repeated, it would be essential to involve more young people in Phase 1, and to attempt to involve them in Phase 2 also.

5.7 Summary

When making shared clinical decisions, parents and clinicians wish to know whether an intervention is effective, and can confer benefits on a child’s ability to live a normal life. Outcomes in clinical trials, which generally measure short-term clinical and physiological efficacy, may not be sufficient to fully evaluate the effects of interventions in ways that are meaningful in clinical practice.

A core set of outcomes for childhood asthma, that not only suits the needs of researchers, but also improves the usefulness of clinical trials from the perspective of clinicians, parents, and policy-makers, can be based on our findings, in conjunction with the recommendations of the recent ATS/ERS taskforce. Agreement amongst a wider group of people involved in such trials should focus on identifying the best ways to measure symptoms, exacerbations, effects of asthma on daily life and adverse effects of therapies. Further discussion should also address the particular questions of whether outcomes reflecting long-term beneficial and harmful effects of treatments should be measured in all trials and, if so, how this should be
done, and which lung function tests, if any, should be included in the core set.

The main implication for the design of core outcome sets is that patients and practising clinicians should be asked which outcomes they feel are important in routine clinical practice, rather than focussing primarily on discussing outcomes which are traditionally measured in clinical trials.

With regard to how this study can be taken forward, there is still much methodological research around the best way to design and implement core outcome sets, and further work to help identify the most appropriate outcomes, and the best way to measure them, for clinical trials in children with asthma. These future directions are discussed in Chapter 6.
Chapter 6

Conclusions and Future directions

6.1 Main conclusions

Several important conclusions can be drawn from the work contained within this thesis. In Chapter 2, in which we describe a systematic review of studies that aim to determine which outcomes to measure in clinical trials in children, we found few relevant paediatric initiatives, and we found that no core outcome set had been agreed for clinical trials in childhood asthma. Most of the existing studies did not use structured consensus techniques to agree on which outcomes to measure in clinical trials. Parents and children were rarely involved in identifying outcomes of importance to them.

In the systematic review to determine which outcomes are measured in clinical trials of regular therapies for childhood asthma, described in Chapter 3, we found that authors frequently evaluate the effects of interventions on short-term disease activity, especially symptoms and physiological markers of lung function, but quality of life, functional status, and long-term beneficial and harmful effects of treatments are rarely assessed.

The main recommendation from this systematic review was that a core outcome set, incorporating outcomes that are important in childhood asthma, should be developed. The Delphi technique was selected as the most
appropriate structured method of reaching consensus, because it enables the involvement of a large number of participants. We conducted a critique of studies that used the Delphi technique for developing core outcome sets, as there was little guidance for researchers who wished to utilise it for this purpose. Between the 15 included studies, we found that the methodology differed considerably, and the quality of reporting of key aspects was variable. Researchers dominated participation in such studies, and patients and families were seldom involved. This review was described in Chapter 4.

Taking into consideration our recommendations from this critique of the Delphi process, as a method of developing core outcome sets, we designed a study that aimed to identify the most important outcomes in childhood asthma. This study was described in Chapter 5. The main methodological approach to this study was to focus on identifying outcomes of particular importance when shared decisions about asthma therapies are made in clinic consultations. In order to do this, we avoided aspects of methodology that are prone to bias, identified in the critique described in Chapter 4. We therefore involved clinicians, parents and young people, rather than researchers or trialists. We also asked open questions to identify potentially important outcomes, rather than presenting participants with a list of outcomes that had been measured in previous clinical trials. We found that the most important outcomes, in clinical practice, were daytime symptoms, nocturnal symptoms, QoL, and mortality. We recommend that these outcomes should be measured, and reported, in all clinical trials of regular, preventative therapies for children with asthma.
The systematic reviews described in Chapters 2 and 4 have raised important questions about core outcome sets in general, which should be addressed before a final core outcome set for childhood asthma is agreed and implemented by those people involved in designing and conducting clinical trials in this condition. These questions relating to how core outcome sets should be developed, disseminated, and implemented, are discussed in Sections 6.2 and 6.3.

6.2 Identifying the best methods for designing core outcome sets

People who wish to develop core outcome sets must make important methodological choices that can affect the validity of their recommendations. As we identified in Chapters 2 and 4, groups have approached the problem of agreeing which outcomes to measure in clinical trials using different methodological techniques. In Chapter 4, we offered guidance that should help inform some of these decisions, but further research should be conducted to improve the quality of core outcome sets.

Such work is important in order to identify attributes of core outcome sets that make them more likely to be acceptable to trialists. If a core outcome set is to be implemented in all relevant studies, it should be accepted and endorsed by all groups of people involved in designing, and reporting, clinical trials. Currently, there is no definition of a minimum set of standards
that helps determine whether a core outcome set has been designed in a rigorous, robust fashion, and should hence be measured, and reported, in all clinical trials in a given condition. In order to identify these attributes, the views of several groups of individuals should be ascertained.

Firstly, there should be discussions and interviews with trialists, to see what would make a core outcome set acceptable from their perspective. Research such as this could be conducted by interviewing trialists who are currently, or recently have been, involved in designing or conducting trials. Open questions could identify whether or not these trialists would consider using a core outcome set, if one were available. Similarly, if a core outcome set has been developed, it would be informative to know, if it were not used, the reasons behind this decision. These interviews would also be helpful in identifying barriers to the implementation of core outcome sets by trialists, such as whether or not core outcome sets are perceived to be too prescriptive.

Organisations who fund clinical trials also have an important role in the promotion and endorsement of core outcome sets. In clinical trials that they fund, they could require that a core outcome set is measured, and may be able to influence trialists towards this. Another group who could be key to aiding the development and utilisation of core outcome sets are drug regulatory agencies, because they directly influence the methodology of clinical trials. In order for these groups to be confident that a core outcome set should be endorsed, they may require that certain criteria are met when it was developed. It would be useful to identify these criteria, so that people
can be given guidance on how to design a core outcome set that will be more likely to be implemented.

As well as interviewing the key personnel outlined above, there may be a different approach to determining the best way to develop core outcome sets. A retrospective audit of clinical trials in which a core outcome set has been developed could be conducted. This should include trials conducted after the core outcome set was published, in order to identify whether or not it was implemented in subsequent research studies. Analysis of the characteristics of these core outcome sets may provide useful information if those which appear to have been implemented are compared against others which have not. For example, there may be differences in the composition of the groups, the methodological techniques, the types of outcomes recommended, or even the journal in which the core outcome set was published.

**Methodological questions relating to the design of core outcome sets**

One relevant problem is whether or not a core outcome set can claim to include all appropriate outcomes, if patients have not been asked which outcomes they feel are important. Further research is required to identify the most appropriate method of involving parents and children in developing core outcome sets, and how their views should influence the final decision about which outcomes to measure in clinical trials. Although we used questionnaires, distributed in out-patient clinics, other groups have
conducted focus groups in patients with fibromyalgia (Mease et al 2008) and degenerative ataxia (Serrano-Aguilar et al 2009). In these conditions, our methodology could be replicated, and the results compared with those of the previous studies. Advantages and disadvantages of these approaches could then be evaluated and compared.

Another methodological problem relates to whether recommendations of a group who reached agreement after open group discussion, rather than using structured consensus techniques, should be considered valid or not. Although we have advocated the use of structured consensus techniques for developing core outcome sets, there have been no studies that empirically compare open group discussion with more structured consensus techniques, for the purpose of developing core outcome sets. Important outcomes identified using these techniques could also be directly compared, and the relative strengths and weaknesses of the methods evaluated.

6.3 Dissemination and implementation of core outcome sets: The COMET initiative

One problem that we encountered throughout the work conducted in this thesis was identifying studies that aimed to determine which outcomes to measure in clinical trials. This is an important barrier to the implementation of core outcome sets. If trialists are not aware of core outcome sets, or do not have access to the study report in which they are described, their usefulness is very much impaired. The recently formed COMET initiative
(Core Outcome Measures in Effectiveness Trials) will hopefully improve
the development and implementation of core outcome sets in clinical
research studies (http://www.liv.ac.uk/nwhtmr/comet/comet.htm - accessed
July 2010). This group advocates the use of core outcome sets, and aims to
bring together individuals who are interested in the standardisation of
outcomes, to collate relevant resources, and foster research in this area, and
help disseminate core outcome sets that have been developed.

One step in the identification of all relevant studies is to repeat the
systematic review, of studies in which outcomes were selected for paediatric
clinical trials, including conditions in all age groups. All relevant studies,
from both systematic reviews, should be compiled, in a searchable database.
This would mean that someone searching for a core outcome set, in a given
condition, would be able to find it, if it existed. This makes it more likely
that existing core outcome sets will be used, and that conditions in which
core outcome sets do not exist could be identified.

Such a database would need to be regularly updated. People involved in the
design of core outcome sets can help the dissemination of their work by
using standardised terminology and keywords in the abstract of the final
publication, so it can be identified easily. These words should be determined
using experts in literature searching and MeSH keyword assignment, such
as librarians. One approach would be that the abstract should mention the
terms “core set” or “standardised outcome set”, and a variety of synonyms
for these terms. The abstract should also clearly define the methodology
used to develop the core outcome set. Journal editors can also play a crucial
role by publishing studies that develop core outcome sets in open-access format.

6.4 A core outcome set for paediatric asthma

In this thesis, we have discussed difficulties associated with the selection and measurement of outcomes for clinical trials of regular therapies for children with asthma, and we recommended that a core outcome set should be agreed for this condition. In Chapter 5 we identified that the most important outcomes in clinical practice are symptoms, exacerbations, mortality, and the effects of asthma on the ability to live a normal daily life.

There is still, however, some work to be done before a final core outcome set is agreed for clinical trials in childhood asthma. The next step is to formally reach as universal a consensus as possible amongst key personnel involved in clinical trials in children with asthma. As childhood asthma a global health problem, this should be an international collaboration. This should be agreed amongst people who design and conduct clinical trials in asthma, and also people who use the results to make decisions about treatments for asthma. This formalisation and universal acceptance of a core outcome set should substantially improve the likelihood that it will be implemented in all relevant trials. An outline for the possible future direction of this work is discussed below.
Who should be involved in agreeing a final core outcome set for clinical trials in children with asthma

It is crucial that the group involved in the final consensus should include clinical trialists, as they will be expected to use the core outcome set in their studies. They would be able to offer useful insights into problems that may arise when measuring and reporting these outcomes. The core outcome set needs to be comprehensive enough to include all of the most relevant outcomes, but it should not be overly prescriptive, to the extent that trialists are discouraged from conducting trials in the first place. Their perspective on an appropriate balance between these two requirements would be very informative.

It may also be appropriate to involve scientific representatives from pharmaceutical companies who are likely to conduct trials of therapies for children with asthma, as it may be more likely that the core outcome set would be properly implemented if these companies agreed that the core set was acceptable. It could also be appropriate to involve representatives from drug regulatory authorities, because their guidance with regard to the evidence required for licensing and marketing purposes largely determines the outcomes measured in trials conducted by the pharmaceutical industry. Regardless of whether representatives from the pharmaceutical industry are included in all discussions about which outcomes to measure, and given equal opportunities as other groups to determine the core outcome set, or whether they observe the process, but do not necessarily have much opportunity to help determine the core outcome set, it is important that their
potential conflicts of interest are recognised, and that any disagreements they may have with the views of other groups are justified.

We have strongly advocated the involvement of those people who use the results of clinical research studies, to make shared decisions about treatment, when developing core outcome sets. In the case of paediatric asthma, we have involved clinicians, parents, and young people. Representatives from these groups should also be involved in the final agreement of the core outcome set. It is not necessary to involve large numbers of these groups, because we have already ascertained the views of several clinicians and young people. It would be appropriate, however, to have representatives from these groups, to ensure that the results from our study are considered when the consensus about the final core outcome set is reached. Although we would hope that people considering our results would agree that the important outcomes should be included in the core set, it is possible that, unless the views of parents and clinicians are strongly advocated, they may be overlooked.

Another group who would benefit from the development of core outcome sets for clinical trials are systematic reviewers, such as those working within the Airways Group of the Cochrane Collaboration. The standardised selection of outcomes across clinical trials would improve the quality of recommendations from Cochrane reviews, and representatives from the Airways group may be interested in helping to develop and agree a final core outcome set in asthma. Their experiences of attempting to meta-analyse the results of clinical trials in asthma may highlight problems that we have
not identified in our work so far. Their involvement may also help to agree the core outcomes that should be considered in Cochrane reviews of interventions for childhood asthma, and included in the Summary of Findings tables.

**How these groups should reach consensus about a core outcome set for childhood asthma**

One way of reaching consensus amongst these individuals is to conduct a Delphi process, via email, and then convene the participants, at a meeting, to discuss issues that have arisen, and to formally agree on the final core outcome set.

In the first round of the Delphi process, participants could be shown the list of outcomes included in the Phase 2 questionnaire we compiled in Chapter 5. They would be asked, for each of these outcomes, whether or not they feel it should be included in the final core set. To help participants, we would distribute the results of the work we described in Chapter 5, and a copy of the ATS/ERS core outcome set. Participants would also be asked to justify their choices, especially for those whose inclusion, or exclusion, from the core outcome set was a matter they felt strongly about. In the second round of the Delphi process we could remove outcomes that the group clearly felt should not be included in the core outcome set, present the overall opinion for the remainder, and a summary of comments from participants, and ask participants whether they would wish to reconsider
their answers. The purpose of this round would be to enable the participants to reiterate their answer, or change their mind. It would also enable us to identify any discrepancies between the various groups of individuals. The third round of the Delphi process could be conducted at a meeting. Any contentious outcomes could be discussed at this meeting, before the core outcome set is agreed at the final vote.

Although we have highlighted certain outcomes as being particularly important, there are others that may also be appropriate for a core outcome set. Whether or not measures of lung function should be measured and reported, in all clinical trials in children with asthma, warrants further discussion by the group agreeing the final core outcome set. Although we have identified that it is not particularly relevant when making decisions in clinical practice, the ATS/ERS taskforce have recommended that it should be measured, and these tests have been established, for several decades, as the most objective measurement of airway obstruction in clinical trials in asthma. Long-term beneficial and harmful effects of interventions also warrant further discussion when the final core set is agreed. We have suggested that these outcomes, which are very infrequently measured in clinical trials, are important to clinicians and parents. Although the difficulties of measuring these outcomes may preclude their measurement in all clinical trials, it may still be appropriate to suggest that, where possible, trialists consider them. It would be important to ascertain the views of clinicians and researchers about this recommendation.
Determining how the outcomes should be measured

In this thesis, we have mainly focussed on identifying which outcomes to measure in clinical trials in asthma. There is also need for a consensus, around how to measure them. In order for this consensus to be evidence-based, research should address which tools are most appropriate for measuring important outcomes. Such work should begin with a systematic review of other studies that have assessed the validity, reliability, or responsiveness of tools for measuring the relevant outcomes, in childhood asthma. If necessary, a prospective comparison of the attributes of these tools should be conducted, amongst large groups of patients in clinical practice, or ongoing clinical trials, to identify which is most appropriate. This work could be conducted between the second and third rounds of the Delphi process described above, and could inform consensus about how the important outcomes should be measured.

Important questions relate to the most appropriate ways of measuring symptoms and QoL. The validity, reliability, responsiveness, and discriminative abilities of these tools should be assessed. There is also a need to determine the most appropriate way to analyse and report these outcomes. For example, they could be reported as mean daily scores, weekly scores, as the proportion of days in which the outcome was worse than a pre-determined cut-off.

Another important issue would be to identify the most acceptable and useful definition for what constitutes an exacerbation. We demonstrated in Chapter 3 that a variety of definitions were used in clinical trials. Exacerbations may
be considered in various ways, such as physiological terms (e.g., reduced FEV1 with increased respiratory rate), health resource utilisation (e.g., hospital attendance), need for treatment (e.g., oral prednisolone), or patient reported events (e.g., “I felt like I had an exacerbation”). There are various ways in which endpoints such as exacerbation can be defined. One is to write various ‘paper patient’ profiles based on the different dimensions of the definition described above. These could be showed to clinicians, and parents, who could be asked whether they would classify the patient as having had an exacerbation. Analysis of the factors that were associated with the most popular definitions of exacerbation could help determine the most useful way of classifying this endpoint.

**Dissemination of the final core set**

Once the final core outcome set has been agreed, it should be disseminated in an easily accessible, open access format. This should be in an open-access journal, and the work should be deposited in the aforementioned registry held by the COMET group. The final core set should also be presented at large international meetings, such as the annual scientific meetings held by the ATS and ERS.
6.5 Final summary

It is crucial that trialists measure the right outcomes. Core outcome sets make it more likely that the most appropriate outcomes are measured, in the right way, by all relevant studies. It is encouraging that important groups are now realising the potential advantages of standardising outcomes across clinical trials. There is still a long way to go before core outcome sets are widely accepted, however. A concerted effort is required, amongst all those involved in clinical research, to encourage the selection of the most appropriate outcomes in clinical trials, and to promote the use of core outcome sets. Better outcomes in trials leads to more useful evidence, which should in turn improve patient care.
List of references


Dalkey, N, 1969, The Delphi method: An experimental study of group opinion


Gubitz, G., Counsell, C., Sandercock, P. & Signorini, D., 2000, Anticoagulants for acute ischaemic stroke, *Cochrane database of systematic reviews (Online)* (2).


Higgins, J.P.T. & Green, S.E., 2008, Cochrane handbook for systematic reviews of interventions Version 5.0. 0 [updated February 2008], *The Cochrane Collaboration*.


Lux, A.L. & Osborne, J.P., 2004, A proposal for case definitions and outcome measures in studies of infantile spasms and West syndrome:


Medical Research Council, 1944, Clinical trial of Patulin in the common cold, Lancet, pp. 373-5.

Medical Research Council, 1948, Streptomycin treatment for Pulmonary Tuberculosis, BMJ (Clinical research ed.), pp. 769-82.


Murphy, M., Black, N. & Lamping, D., 1998, Consensus development methods, and their use in clinical guideline development, .


Rieger, W.G., 1986, Directions in Delphi developments: dissertations and


Temple, R.J., 1995, 1 A regulatory authority's opinion about surrogate endpoints, *Clinical measurement in drug evaluation*, p. 1.


Wilkins, T., Gillies, R.A. & Davies, K., 2005, EMBASE versus MEDLINE for family medicine searches: can MEDLINE searches find the forest or a tree? Canadian Family Physician, 51(6), p. 848.


Appendices
Appendix 1 – Search strategies used in the systematic review of studies that aimed to determine which outcomes to measure in clinical trials in children (Chapter 2)

MEDLINE (1950-2006) and CINAHL (1982-2006) were searched in January 2007 using the following search strategy:

1  (clinical trials or clinical trials).mp.
2  Clinical trials, phase III.mp.
3  Clinical trials, phase IV/
4  multicentre studies.mp. or multicenter studies/
5  controlled clinical trials.mp. or controlled clinical trials/
6  randomized controlled trial.mp. or randomized controlled trial/
7  randomised controlled trial.mp. or randomised controlled trial/
8  randomized controlled trials.mp. or randomized controlled trials/
9  randomised controlled trials.mp. or randomised controlled trials/
10 controlled clinical trials.mp. or controlled clinical trials/
11  (Clinical adj3 trial$).tw.
12  (controlled adj3 trial$).tw.
13  rcts.mp.
14  trial$.mp.
15  therapeutic human experimentation.mp. or therapeutic human experimentation/
16  1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15
17  methodology, research.mp.
18  research methodology.mp.
19  methodology.mp.
20  research techniques.mp.
21  research technics.mp.
22  research design.mp. [mp=ti, hw, ab, it, ot, nm]
23  experimental design.mp. [mp=ti, hw, ab, it, ot, nm]
24  methodological stud$.mp. [mp=ti, hw, ab, it, ot, nm]
25  design, experimental.mp. [mp=ti, hw, ab, it, ot, nm]
26  study, methodological.mp. [mp=ti, hw, ab, it, ot, nm]
27  studies, methodological.mp.
28  group processes.mp. or group processes/
29  consensus.mp. or consensus/
30  consensus development conferences, NIH/ or "consensus development conference NIH".mp. [mp=ti, hw, ab, it, ot, nm]
31  consensus development conferences/ or "consensus development conference".mp. [mp=ti, hw, ab, it, ot, nm]
SCOPUS was searched in January 2007 using the following search strategy:

```
((TITLE-ABS-KEY(clinical trial*)) OR (TITLE-ABS-KEY(controlled clinical trial*)) OR (TITLE-ABS-KEY(randomized controlled trial*)) OR (TITLE-ABS-KEY(outcome assessment (health care)) OR (Outcome Assessment (Health Care)) OR (treatment outcome mp. OR treatment outcome)) OR (outcome measures.mp. OR outcome studies.mp. OR outcomes assessment.mp. OR assessment, outcomes.mp. OR treatment failure.mp. OR treatment failure)) OR (outcome$.mp. OR outcome determination/ OR outcome.mp. OR endpoint$.mp. OR end point$.mp. OR 42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 51 or 52 or 53) OR (child$.mp. OR child$ OR paediatric$.mp. OR pediatric$.mp. OR infant/ OR infant$.mp. OR young$.mp. OR Toddler.mp. OR bab$.mp. OR child, preschool/ OR preschool$.mp. OR Pre-school.mp. OR Adolesc$.mp. OR adolescent/ OR Teenage$.mp. OR youth$.mp. OR 55 or 56 or 57 or 58 or 59 or 60 or 61 or 62 or 63 or 64 or 65 or 66 OR 16 and 41 and 54 and 67) OR (remove duplicates from 68)
```

"congresses [publication type]" OR congresses OR congresses.mp.
consumer participation.mp. OR consumer participation/
patient participation.mp. OR patient participation/
literature review.mp. OR "Review Literature"/
systematic review.mp.
17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40
"outcome assessment (health care)".mp. OR "Outcome Assessment (Health Care)"/
treatment outcome.mp. OR treatment outcome/
patient outcome assessment.mp.
outcome measures.mp.
outcome studies.mp.
outcomes assessment.mp.
assessment, outcomes.mp.
treatment failure.mp. OR treatment failure/
outcome$.mp.
outcome determination/ OR outcome.mp.
eendpoint$.mp.
end point$.mp.
42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 51 or 52 or 53
child$.mp. OR child$ OR paediatric$.mp. OR pediatric$.mp.
infant/ OR infant$.mp.
young$.mp.
Toddler.mp.
bab$.mp.
child, preschool/ OR preschool$.mp.
Pre-school.mp.
Adolesc$.mp. OR adolescent/
Teenage$.mp.
youth$.mp.
55 or 56 or 57 or 58 or 59 or 60 or 61 or 62 or 63 or 64 or 65 or 66
16 and 41 and 54 and 67
remove duplicates from 68
Cochrane database of systematic reviews, Cochrane database of abstracts reviews of effects, Cochrane database of methodology reviews, Cochrane methodology register and Cochrane central register of controlled trials (CENTRAL) were searched in December 2006 using the following search strategy:

#1 clinical trial*:ti,ab,kw OR controlled clinical trial:ti,ab,kw
#2 randomi?ed controlled trial*:ti,ab,kw
#3 multicentre stud* OR multicenter stud*
#4 (#1 OR #2 OR #3)
#5 research methodol*:ti,ab,kw OR methodol:ti,ab,kw
#6 research design:ti,ab,kw OR experimental design:ti,ab,kw
#7 methodolog* stud*:ti,ab,kw
#8 consensus:ti,ab,kw OR consensus development:ti,ab,kw
#9 focus group*:ti,ab,kw
#10 group processes:ti,ab,kw
#11 consumer participation:ti,ab,kw OR patient participation:ti,ab,kw
#12 literature review:ti,ab,kw OR systematic review:ti,ab,kw
#13 (#5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12)
#14  outcome*:ti,ab,kw
#15  outcome*:ti,ab,kw
#16  outcome measure*:ti,ab,kw
#17  outcome stud*:ti,ab,kw OR outcome assessment*:ti,ab,kw
#18  patient outcome:ti,ab,kw OR treatment outcome:ti,ab,kw
#19  (#14 OR #15 OR #16 OR #17 OR #18)
#20  paediatric*:ti,ab,kw OR pediatric*:ti,ab,kw
#21  child*:ti,ab,kw
#22  bab*:ti,ab,kw
#23  infan*:ti,ab,kw
#24  toddler*:ti,ab,kw
#25  preschool:ti,ab,kw OR pre-school:ti,ab,kw
#26  adolescen*:ti,ab,kw
#27  youth*:ti,ab,kw
#28  teenage*ti,ab,kw
#29  (#20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #28)
#30  (#4 AND #13 AND #19 AND #29)
## Appendix 2 – Studies excluded from the review described in Chapter 2

<table>
<thead>
<tr>
<th>Study</th>
<th>Reason for exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adams 1998 [1]</td>
<td>This paper described the rationale behind the design of a clinical trial investigating stroke prevention in children with sickle cell anaemia. The paper did not, however, develop or use methodology for selecting the outcomes used in the trials.</td>
</tr>
<tr>
<td>Allen 1985[2]</td>
<td>This review article discussing design issues in clinical trials of children with brain tumours is based on the author's personal opinion. Although features of some outcomes are discussed, this does not qualify as methodology of selecting outcomes.</td>
</tr>
<tr>
<td>Andrasik 2005[3]</td>
<td>This review article discussing design issues in clinical trials of children with headache is based on the author's personal opinion. Although features of some outcomes are discussed, methodology of selecting outcomes is not used or developed.</td>
</tr>
<tr>
<td>Arnold 2003[4]</td>
<td>This study validated the use of parent-reported outcomes in clinical trials of children with autism. The study was excluded, however, because it did not relate to the actual methodology of selecting an outcome, but the validation of an outcome which was already selected.</td>
</tr>
<tr>
<td>Ball 2003[5]</td>
<td>This study aimed to develop a definition of what constitutes a 'responder' in clinical trials of children with abdominal pain. Therefore the authors used methodology to refine an existing outcome, rather than using methodology for selecting outcomes for use in clinical trials of children with abdominal pain.</td>
</tr>
<tr>
<td>Beghi 1993[6]</td>
<td>This paper describes the rationale behind the design of a clinical trial, including the outcomes which were selected. The paper did not, however, describe methodology for selecting outcomes for use in clinical trials.</td>
</tr>
<tr>
<td>Bower 1994[7]</td>
<td>The review article discussed in detail the potential ways of measuring outcomes, but did not describe the selection of outcomes for use in clinical trials of children with cerebral palsy.</td>
</tr>
<tr>
<td>Children’s Amalgam Trial Study Group [8]</td>
<td>This paper described the rationale behind various aspects of a clinical trial in children requiring dental restoration. The authors do not describe the methodology behind selecting the outcome.</td>
</tr>
<tr>
<td>Author</td>
<td>Year</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>Curley</td>
<td>2005[9]</td>
</tr>
<tr>
<td>Curley</td>
<td>2006[10]</td>
</tr>
<tr>
<td>Guilbert</td>
<td>2004[12]</td>
</tr>
<tr>
<td>Hargrave</td>
<td>2006[13]</td>
</tr>
<tr>
<td>Hollander</td>
<td>2004[14]</td>
</tr>
<tr>
<td>Juniper</td>
<td>2001[15]</td>
</tr>
<tr>
<td>King</td>
<td>2004[16]</td>
</tr>
<tr>
<td>Kirkham</td>
<td>2004[17]</td>
</tr>
<tr>
<td>Law</td>
<td>1989[18]</td>
</tr>
<tr>
<td>Mahoney</td>
<td>2006[19]</td>
</tr>
<tr>
<td>Marshall</td>
<td>2005[20]</td>
</tr>
<tr>
<td>Reference</td>
<td>Year</td>
</tr>
<tr>
<td>-----------</td>
<td>------</td>
</tr>
<tr>
<td>Martin 2006[21]</td>
<td></td>
</tr>
<tr>
<td>McBride 1990 [22]</td>
<td></td>
</tr>
<tr>
<td>McDaid 2005[23]</td>
<td></td>
</tr>
<tr>
<td>Mehta 2003 [24]</td>
<td></td>
</tr>
<tr>
<td>Meltzer 2006[25]</td>
<td></td>
</tr>
<tr>
<td>Mohanraj 2003 [26]</td>
<td></td>
</tr>
<tr>
<td>Ozolins 2005 [27]</td>
<td></td>
</tr>
<tr>
<td>Pauwels 2001 [28]</td>
<td></td>
</tr>
<tr>
<td>Ravelli 1997 [29]</td>
<td></td>
</tr>
<tr>
<td>Romagnoli 2001 [30]</td>
<td></td>
</tr>
<tr>
<td>Rosenfeld 2001 [31]</td>
<td></td>
</tr>
<tr>
<td>Author</td>
<td>Year</td>
</tr>
<tr>
<td>--------------</td>
<td>------</td>
</tr>
<tr>
<td>Rudorfer</td>
<td>1993</td>
</tr>
<tr>
<td>Ruperto</td>
<td>1996</td>
</tr>
<tr>
<td>Schoenfeld</td>
<td>2002</td>
</tr>
<tr>
<td>Schonbeck</td>
<td>2005</td>
</tr>
<tr>
<td>Shapiro</td>
<td>1999</td>
</tr>
<tr>
<td>Sharek</td>
<td>2002</td>
</tr>
<tr>
<td>Spahn</td>
<td>2003</td>
</tr>
<tr>
<td>Spencer</td>
<td>1997</td>
</tr>
<tr>
<td>Thase</td>
<td>1999</td>
</tr>
<tr>
<td>Theander</td>
<td>2005</td>
</tr>
<tr>
<td>Van Zanten</td>
<td>2006</td>
</tr>
<tr>
<td>Walsh</td>
<td>2003</td>
</tr>
<tr>
<td>Study</td>
<td>Reason for exclusion</td>
</tr>
<tr>
<td>-------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Walsh 2004[44]</td>
<td>In this study, the authors assess the effect of a physiological definition on the incidence of bronchopulmonary dysplasia. The study does not select outcomes for use in future clinical trials.</td>
</tr>
<tr>
<td>Zhang 2000[45]</td>
<td>This study assesses the performance characteristics of outcomes used in neonatal clinical trials, but is not about selection of outcome domains or outcomes.</td>
</tr>
<tr>
<td><strong>Studies suggested by MCRN CSG members</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Study</strong></td>
<td><strong>Reason for exclusion</strong></td>
</tr>
<tr>
<td>Aurora 2005[46]</td>
<td>In this study the correlation of various tests of lung function in children with CF was assessed. The study did not relate to the selection of which outcomes should be measured in clinical trials.</td>
</tr>
<tr>
<td>Brody 2004[47]</td>
<td>In this study the authors assessed the correlation between high resolution CT scanning and pulmonary function tests in children with CF. The authors do not use methodology to determine which outcomes should actually be measured in clinical trials of children with CF however.</td>
</tr>
<tr>
<td>Clayton 2002[48]</td>
<td>Consensus about management of children with 21-hyroxylation deficiency, not relating to clinical trial design</td>
</tr>
<tr>
<td>Dunger 2004[50]</td>
<td>Consensus statement about management of children with Diabetic Ketoacidosis, not relating to clinical trial design</td>
</tr>
<tr>
<td>Hughes 2006[51]</td>
<td>Consensus statement about children suffering from intersex disorders, not relating to clinical trial design.</td>
</tr>
<tr>
<td>Lum 2007[52]</td>
<td>In this study the authors assessed the extent at which multiple breath washout parameters are able to detect airway dysfunction in children with CF. The study did not relate to the selection of outcomes for use in clinical trials.</td>
</tr>
<tr>
<td>Ordonez 2003[53]</td>
<td>This study assesses the performance characteristics of inflammatory and microbiological markers in clinical trials of children with cystic fibrosis. The authors do not use methodology to determine which outcomes should actually be measured in these clinical trials however.</td>
</tr>
<tr>
<td>Ordonez 2004[54]</td>
<td>In this study the authors assess the reproducibility of inflammatory markers as an outcome measure in children with CF. The authors do not use methodology to determine which outcomes should actually be measured in clinical trials of children with CF however.</td>
</tr>
<tr>
<td>Ratjen 2006[55]</td>
<td>In this review article the authors discuss some ways of measuring airway inflammation in children with CF. They do not use methodology however to determine which outcomes should be measured in clinical trials if children with this condition.</td>
</tr>
</tbody>
</table>
In this study the authors assess the safety and usefulness of inducing sputum from children with CF in order to measure various inflammatory and microbiological markers that could be used as outcomes in clinical trials. The study does not use methodology to determine which outcomes should actually be measured in clinical trials of children with CF.

In this study consumers were involved in the design of a clinical trial of children with migraine. The consumers were not involved in the selection of which outcomes should be measured in the trial (personal communication from Dr Whitehouse).

References to studies excluded in the systematic review described in Chapter 2
Appendix 3 - Description of studies included in the systematic review described in Chapter 2

<table>
<thead>
<tr>
<th>Study</th>
<th>Condition</th>
<th>Type of Study</th>
<th>People involved selecting outcomes</th>
<th>Patient involvement</th>
<th>Methods used to select outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goldstein 2005 [1]</td>
<td>Sepsis</td>
<td>Consensus statement</td>
<td>Paediatric critical care specialists; physicians and scientists with research experience in paediatric sepsis, members of past consensus conferences on adult sepsis</td>
<td>No</td>
<td>The group arrived at consensus by way of semi-structured discussion.</td>
</tr>
<tr>
<td>De Rouen 2002 [2]</td>
<td>Dental restoration</td>
<td>Paper describes the design of specific trial– the authors of the trial consulted with a multidisciplinary team of experts to select the outcomes measured in the trial. (No of participants not reported)</td>
<td>Physicians, epidemiologists, biostatisticians. Number of participants not stated</td>
<td>No</td>
<td>The consensus was reached by discussion within the group.</td>
</tr>
<tr>
<td>Griffiths 2005 [3]</td>
<td>Crohn’s disease</td>
<td>Consensus statement</td>
<td>Clinicians with expertise in paediatric inflammatory bowel disease; clinical trial design specialists (Number of participants not reported)</td>
<td>No</td>
<td>Semi-structured discussion was used to select the most appropriate outcome domains and outcomes to measure</td>
</tr>
<tr>
<td>Pavletic 2006 [4]</td>
<td>Graft Versus Host disease (GVHD)</td>
<td>Consensus statement regarding trials of children with GVHD, including formulation of a preliminary core set of outcomes.</td>
<td>Clinical experts in GVHD; Experts from other fields (gastroenterology and rheumatology) who were experienced in selecting outcomes for use in clinical trials within those fields. (Number of participants not stated)</td>
<td>Yes – at the end of the report, the authors acknowledge “patients and patient and research advocacy groups”. The level of their involvement was not reported by the authors.</td>
<td>The group arrived at consensus by way of semi-structured discussion.</td>
</tr>
<tr>
<td>Finer 2006 [5] On behalf of NDDI</td>
<td>Neonatal apnoea</td>
<td>Consensus statement</td>
<td>Physicians, research experts (Number of participants not reported)</td>
<td>No</td>
<td>The consensus was reached by semi-structured discussions by the group.</td>
</tr>
<tr>
<td>Study</td>
<td>Condition</td>
<td>Type of Study</td>
<td>People involved in selecting outcomes</td>
<td>Patient involvement</td>
<td>Methods used to select outcome</td>
</tr>
<tr>
<td>-------</td>
<td>-----------</td>
<td>---------------</td>
<td>--------------------------------------</td>
<td>---------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>Short 2006 [6] On behalf of NDDI</td>
<td>Neonatal cardiovascular instability</td>
<td>Consensus conference regarding trials of neonates with cardiovascular instability</td>
<td>Physicians, research experts (Number of participants not reported)</td>
<td>No</td>
<td>The group arrived at consensus by using semi-structured discussion.</td>
</tr>
<tr>
<td>Anand 2005 Anand 2006 [7,8] on behalf of the NDDI</td>
<td>Neonatal pain</td>
<td>1. Systematic review of literature relating to trial design in studies of neonatal pain 2. Consensus statement regarding trial design considerations relevant to clinical trials of pain relief for neonates</td>
<td>Experts on neonatal pain, clinical researchers (Number of participants not reported)</td>
<td>No</td>
<td>A systematic review of literature relating to trial design in studies of neonatal pain was performed, and specific articles relating to outcome measurement were reviewed. These findings were discussed amongst a group of experts, using a semi-structured discussion approach, and a consensus statement regarding the outcomes that should be measured was developed. Although a lot of the discussion centred around tools to measure outcomes, there was also discussion about composite outcomes, and also about outcome domains, such as long-term neurodevelopmental outcome.</td>
</tr>
<tr>
<td>Roth 2006[9] On behalf of NDDI</td>
<td>Neonatal Postoperative Cardiac Dysfunction</td>
<td>Consensus statement regarding trials of neonates with Postoperative Cardiac Dysfunction</td>
<td>Physicians, research experts. (number of participants not reported)</td>
<td>No</td>
<td>The group arrived at consensus by using semi-structured discussion.</td>
</tr>
<tr>
<td>Clancy 2006 [10] On behalf of NDDI</td>
<td>Neonatal seizures</td>
<td>Consensus statement</td>
<td>Physicians, research experts (Number of participants not reported)</td>
<td>No</td>
<td>The consensus was reached by semi-structured discussions by the group.</td>
</tr>
<tr>
<td>Study</td>
<td>Condition</td>
<td>Type of Study</td>
<td>People involved</td>
<td>Patient involvement</td>
<td>Methods used to select outcome</td>
</tr>
<tr>
<td>-------</td>
<td>-----------</td>
<td>---------------</td>
<td>-----------------</td>
<td>--------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>Lux 2004 [11] Osborne 2001 [12] On behalf of the West Delphi collaboration</td>
<td>Infantile spasms</td>
<td>Consensus statement regarding trials of children with IS, including formulation of a preliminary core set of outcomes.</td>
<td>Authors who had published articles on infantile spasms during the previous 10 years were invited to join the group, and were asked to invite others who may be interested in participating. (Questionnaires were sent to 133 people, and there were eventually 31 participants)</td>
<td>No</td>
<td>Delphi consensus performed over 6 ‘rounds’ of questions: 1) Multiple Choice Questions (MCQ) covering various aspects of clinical trial design, including outcomes. 2) Qualitative comments and MCQ questions regarding the same questions, having fed the results of round 1 back to the group. 3) Formulation of statements from rounds 1 and 2 which represented majority opinion. Participants were invited to respond as to whether they agreed or disagreed with these. 4) These statements were modified, and participants commented on their suitability and content. 5) and 6) consisted of formulation of a draft and subsequently a final paper which were altered according to the group’s comments</td>
</tr>
<tr>
<td>Carlson 2003 [13]</td>
<td>Bipolar affective disorder</td>
<td>Consensus workshop regarding trial design considerations relevant to clinical trials of adolescents with bipolar disorder.</td>
<td>Clinical researchers with expertise in bipolar illness, pharmaceutical industry sponsors, staff of the Food and Drug Administration, representatives of families with affected children. (Total number of participants: 53)</td>
<td>Yes. The group included parents of children who were affected by bipolar affective disorder.</td>
<td>Several separate groups used semi-structured discussion to come to consensus on several issues relating to clinical trial design, and one group was specifically given the remit of discussing outcomes. This included discussion on which primary and secondary outcomes should be measured, and which aspects of mania or depression best represent change in a child’s condition.</td>
</tr>
<tr>
<td>Study</td>
<td>Condition</td>
<td>Type of Study</td>
<td>People involved in selecting outcomes</td>
<td>Patient involvement</td>
<td>Methods used to select outcome</td>
</tr>
<tr>
<td>------------</td>
<td>--------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------</td>
<td>---------------------</td>
<td>-----------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>LaFrance (2006) [14]</td>
<td>Non-Epileptic Seizures (NES)</td>
<td>Consensus statement about clinical trials of patients suffering from NES</td>
<td>Multidisciplinary group of neurologists, psychiatrists, neuropsychiatrists, psychologists, statisticians, nurses and other researchers familiar with NES (44 participants)</td>
<td>No</td>
<td>Structured and focussed discussion. The group discussed the question “which variables or domains should be regarded as reflecting outcome?” The group identified areas of outcomes relevant to NES that could potentially be used as outcomes reflecting the effects of an intervention. These included psychosocial outcomes, clinical outcome, psychiatric status, health-related quality of life, medical resource utilisation, and psychophysiological markers.</td>
</tr>
<tr>
<td>Smith 1996 [15]</td>
<td>Asthma</td>
<td>Questionnaire-based survey of health professionals and researchers.</td>
<td>Researchers and clinical experts with experience in treating asthma. (14 participants)</td>
<td>No</td>
<td>Questionnaires to health professionals and researchers asking which outcomes they would use in various scenarios, including clinical trials, to assess whether a patient’s asthma had been improved by an intervention. The participants were also asked to provide an estimate of how valid they thought the indicators were.</td>
</tr>
<tr>
<td>Ramsey 1994 [16]</td>
<td>Cystic Fibrosis</td>
<td>Consensus statement regarding trials of children with Cystic Fibrosis, including formulation of a preliminary core set of outcomes.</td>
<td>Clinicians with expertise in CF, laboratory and clinical researchers in the field of CF, representatives from the U.S. FDA (Number of participants not reported)</td>
<td>No</td>
<td>The group arrived at consensus by using semi-structured discussion.</td>
</tr>
<tr>
<td>Study</td>
<td>Condition</td>
<td>Type of Study</td>
<td>People involved in selecting outcomes</td>
<td>Patient involvement</td>
<td>Methods used to select outcome</td>
</tr>
<tr>
<td>-------</td>
<td>-----------</td>
<td>---------------</td>
<td>----------------------------------------</td>
<td>---------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>Miller 2001 Rider 2002 Rider 2003 Rider 2004 [17–20] On behalf of the IMACS collaboration</td>
<td>Juvenile idiopathic inflammatory myopathies (IIM)</td>
<td>Development of core set of outcomes for use in trials of children with IIM, by formulation of a consensus statement</td>
<td>Adult and paediatric specialists, patient support group leaders with expertise in IIM. This group was called The International Myositis Outcomes Assessment Collaborative Study Group (IMACS). (Number of participants not reported)</td>
<td>Patient support group leaders were the parents of children with IIM.</td>
<td>It was decided that in order to fully understand the totality of effects of interventions on patients with IIM, the outcomes should measure disease activity, disease damage and quality of life. A literature review was performed to review the performance characteristics of outcome measures used in IIM clinical trials, and Delphi technique was used to determine which outcomes best represented these domains. A definition of improvement was then developed, and this was subsequently validated.</td>
</tr>
<tr>
<td>Oddis 2005 [21] Describing the work of the IMACS collaboration</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>On behalf of the PRINTO collaboration Ruperto 2003 Ruperto 2004 Ruperto 2006 [22–24]</td>
<td>Systemic lupus erythematosus (SLE) and dermatomyositis (DM)</td>
<td>Development of a core set of outcomes for use in trials of children with SLE and DM</td>
<td>Specialists in paediatric rheumatology 267 people were initially involved, followed by a meeting of 40 participants</td>
<td>No</td>
<td>1) Development of a core set of outcomes Phase 1: (Delphi Technique) A questionnaire was sent to the group, and participants were asked to rank the 10 variables they judged as clinically most important when determining whether a patient with SLE or DM has responded to therapy. Variables ranked by at least 10 responders were then listed alphabetically and participants were asked again to rank their top 10 choices. Phase 2: (Nominal group technique) A four-day international consensus conference was attended by 40 participants. Using 5 exercises, the</td>
</tr>
</tbody>
</table>
core set of outcomes was determined: 1) Classify all variables into one of 2 domains (disease activity and disease damage). The participants were invited to suggest any other domains. 2) Classify variables into ‘concepts’ of disease activity and damage 3) Select and rank the domains that should be included in the core set 4) Select the variables that should be used to measure these domains

2) Validation of core outcomes
This was done in clinical practice, on patients who were starting a new medication. Validation of the following characteristics was conducted: feasibility; face and content validity; responsiveness, discriminative ability; convergent construct validity; internal consistency

3) Development of a definition of improvement
<table>
<thead>
<tr>
<th>Condition</th>
<th>Type of Study</th>
<th>People involved in selecting outcomes</th>
<th>Patient involvement</th>
<th>Methods used to select outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Giannini 1997 [25]</td>
<td>Juvenile Arthritis</td>
<td>Development and validation of a set of core outcomes using consensus formation techniques, and development of a definition of improvement for individual patients.</td>
<td>No</td>
<td>1) 25 variables which had been used in juvenile arthritis (JIA) trials were listed in a questionnaire. 2) This questionnaire was sent to a 16-member ‘advisory council’ who were asked to rank their top 6 choices, and list other variables which were not included on the list. 16 variables received votes, and these became the ‘candidate variables’. 3) The performance characteristics (validity, reliability, sensitivity to change, redundancy) of the candidate variables were reviewed using existing literature. 4) Using nominal group technique, the group selected a preliminary core set of 6 outcomes. These were physician global assessment of disease activity, parent or patient reported assessment of overall well being, functional ability, number of joints with active arthritis, number of joints with limited range of movement, and erythrocyte sedimentation rate. 5) International consensus on the acceptability of this core set of variables was ascertained using a questionnaire. 6) The multicollinearity of these outcomes was assessed using real patients.</td>
</tr>
</tbody>
</table>
7) A definition of improvement was developed

Reference List to studies included in Chapter 2


Appendix 4 – Outcomes selected in the studies included in the systematic review described in Chapter 2

<table>
<thead>
<tr>
<th>Speciality</th>
<th>Condition</th>
<th>Study</th>
<th>Disease Activity</th>
<th>Physical consequences of disease – damage/complications</th>
<th>Adverse effects of therapy</th>
<th>Functional status</th>
<th>Social outcome/ family outcome, including Quality of Life</th>
<th>Resource utilisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Critical care:</td>
<td></td>
<td>Sepsis (Goldstein)</td>
<td></td>
<td>Appropriate biomarkers (as secondary or pharmaco-dynamic endpoints)</td>
<td>Organ failure-free days; Mortality (not as the primary outcome)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gastroenterology:</td>
<td>Crohn’s disease (Griffiths)</td>
<td></td>
<td>Acute: Primary – CDAI Secondary: PCDAI Remission: CDAI</td>
<td></td>
<td></td>
<td>Change in Pediatric Overall Performance Category score between admission and discharge; overall functioning at 3 or 6 months</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Physician assessment; Parent/patient assessment; Height velocity</td>
<td>Growth</td>
<td></td>
<td></td>
<td>HRQoL</td>
<td></td>
</tr>
<tr>
<td>Disease Activity</td>
<td>Physical consequences of disease – damage/ complications</td>
<td>Adverse effects of therapy</td>
<td>Functional status</td>
<td>Social outcome/ family outcome, including Quality of Life</td>
<td>Resource utilisation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>---------------------------------------------------------</td>
<td>---------------------------</td>
<td>------------------</td>
<td>--------------------------------------------------------</td>
<td>---------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Haematology:</strong></td>
<td></td>
<td></td>
<td>Clinician assessed: Grip strength; 2-minute walk time; Karnofsky/ Lansky scale</td>
<td></td>
<td>SF-36 CHRIS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>GVHD</strong> (Pavletic)</td>
<td>Signs-organ specific measures Symptoms-clinician reported Symptoms – patient reported Global rating-physician Global rating-patient reported</td>
<td>Associated postnatal morbidity;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Neonatology:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Neonatal apnoea</strong> (Finer)</td>
<td>Frequency/duration/severity of apnoea</td>
<td>Associated postnatal morbidity;</td>
<td>Neurodevelopment at 18 months</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Neonatology:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Neonatal cardio-vascular instability</strong> (Short)</td>
<td>Measures of organ perfusion</td>
<td><strong>Primary outcome:</strong> Combined endpoint of mortality or severe neurological outcome <strong>Secondary outcomes:</strong> NEC;ROP;BPD;</td>
<td>Adverse effects of drugs- arrhythmia, hypertension, seizures, hormonal effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Neonatology:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pain</strong> (Anand)</td>
<td>Acute pain measurements: Behavioural/ Haemodynamic/ Metabolic/ Respiratory/ Renal; Need for supplemental opiates; Frequency of postnatal complications of prematurity</td>
<td>Neurobehavioural assessment</td>
<td>Global developmental delay Intervention-specific adverse effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Discontinued**

- **Disease Activity**

  - **Haematology:**
    - **GVHD** (Pavletic)
      - Signs-organ specific measures
      - Symptoms-clinician reported
      - Symptoms – patient reported
      - Global rating-physician
      - Global rating-patient reported

  - **Neonatology:**
    - **Neonatal apnoea** (Finer)
      - Frequency/duration/severity of apnoea
    - **Neonatal cardio-vascular instability** (Short)
      - Measures of organ perfusion
    - **Pain** (Anand)
      - Acute pain measurements: Behavioural/ Haemodynamic/ Metabolic/ Respiratory/ Renal; Need for supplemental opiates; Frequency of postnatal complications of prematurity

- **Physical consequences of disease – damage/ complications**

  - Associated postnatal morbidity;

- **Adverse effects of therapy**

  - Adverse effects of drugs- arrhythmia, hypertension, seizures, hormonal effects

- **Functional status**

  - Clinician assessed: Grip strength; 2-minute walk time; Karnofsky/ Lansky scale
  - Patient assessed: HAP; ASK;

- **Social outcome/ family outcome, including Quality of Life**

  - SF-36
  - CHRIS

- **Resource utilisation**

  - Duration of hospital stay; Duration of assisted ventilation
  - Number of days hospitalized
  - Length of hospital stay/ ventilation/ NICU stay as represented by physiological definition of criteria to account for inter-centre variations in practice
<table>
<thead>
<tr>
<th>Disease Activity</th>
<th>Physical consequences of disease – damage/ complications</th>
<th>Adverse effects of therapy</th>
<th>Functional status</th>
<th>Social outcome/ family outcome, including Quality of Life</th>
<th>Resource utilisation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Neonatology</strong></td>
<td><strong>Neonatal post-operative cardiac dysfunction</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Roth)</td>
<td>Duration of delayed sternal closure; Mortality beyond 30th postoperative day;</td>
<td>Neurodevelopment at 1 to 2 years</td>
<td></td>
<td>Duration of assisted ventilator; Duration of NICU stay; Duration of hospital stay after cardiac surgery;</td>
</tr>
<tr>
<td><strong>Neonatology:</strong></td>
<td><strong>Seizures</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Clancy)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Neurology:</strong></td>
<td><strong>Infantile spasms</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(West Delphi Group)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Primary: Cessation of EEG-detected seizures</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Clinical outcome –spasm cessation Electroclinical outcome- resolution of hypsarrhythmia Relapse-free response; Continued subtle spasm; Time to relapse;Presence of other seizures</td>
<td>Development at 2 years; Death and other serious adverse effects associated with the illness; Nonserious adverse events associated with the illness; Presence of and progression to other seizure types</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>continued</td>
<td>Disease Activity</td>
<td>Physical consequences of disease – damage/complications</td>
<td>Adverse effects of therapy</td>
<td>Functional status</td>
<td>Social outcome/ family outcome, including Quality of Life</td>
</tr>
<tr>
<td>-----------</td>
<td>-----------------</td>
<td>--------------------------------------------------------</td>
<td>-----------------------------</td>
<td>------------------</td>
<td>---------------------------------------------------------</td>
</tr>
<tr>
<td>Psychiatry: Bipolar affective disorder (Carlson)</td>
<td>Primary: Mania (YMS); Depression (CDRS). Secondary: ADHD symptomology; Aggressive behaviour; Global improvement</td>
<td></td>
<td></td>
<td>Academic outcome and cognitive function</td>
<td>Family outcome-CHQ; Other social effects</td>
</tr>
<tr>
<td>Psychiatry: Non-epileptic seizures (LaFrance)</td>
<td>Seizure frequency</td>
<td>Depression; Personality characteristics; Arousal</td>
<td></td>
<td></td>
<td>HRQoL, including illness perceptions and individual concerns; Employment; Family functioning</td>
</tr>
<tr>
<td>Respiratory: CF (Ramsey)</td>
<td>&lt; 6 years of age Chest x-ray score; Oxygenation; Inflammatory markers; Illness severity; Bronchoscopy/ BAL/brushings &gt; 6 years of age As above, and Spirometry; Sputum: microbiology/ DNA</td>
<td>Growth; Frequency of pulmonary exacerbations</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disease Activity</td>
<td>Physical consequences of disease – damage/complications</td>
<td>Adverse effects of therapy</td>
<td>Functional status</td>
<td>Social outcome/ family outcome, including Quality of Life</td>
<td>Resource utilisation</td>
</tr>
<tr>
<td>------------------</td>
<td>--------------------------------------------------------</td>
<td>----------------------------</td>
<td>-------------------</td>
<td>--------------------------------------------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Respiratory: Asthma-&lt;br&gt;Acute (Smith)</td>
<td>Symptoms&lt;br&gt;FEV1/FVC&lt;br&gt;PEFR</td>
<td></td>
<td></td>
<td></td>
<td>Admission to hospital</td>
</tr>
<tr>
<td></td>
<td>Symptoms&lt;br&gt;FEV1/FVC&lt;br&gt;PEFR&lt;br&gt;Frequency of medication use;&lt;br&gt;Bronchial hyper-responsiveness</td>
<td></td>
<td>Functional status</td>
<td>HRQoL</td>
<td></td>
</tr>
<tr>
<td>Rheumatology: IIM (IMACS)</td>
<td>Global activity&lt;br&gt;(physician/parent/patient assessment);&lt;br&gt;Muscle strength (Manual testing);&lt;br&gt;Enzymes:&lt;br&gt;CK, aldose, LD, AST, ALT;&lt;br&gt;CMAS;&lt;br&gt;Extra-skeletal muscle disease</td>
<td>Core set:&lt;br&gt;Global damage assessment;&lt;br&gt;Assessment of different organ systems (VAS/SDI)&lt;br&gt;Extended core set:&lt;br&gt;MRI: muscle fibrosis/scarring/atrophy; serum creatinine; cutaneous assessment</td>
<td>HAQ/CHAQ</td>
<td>HRQoL (SF36)</td>
<td></td>
</tr>
<tr>
<td>Disease Activity</td>
<td>Physical consequences of disease – damage/complications</td>
<td>Adverse effects of therapy</td>
<td>Functional status</td>
<td>Social outcome/ family outcome, including Quality of Life</td>
<td>Resource utilisation</td>
</tr>
<tr>
<td>------------------</td>
<td>----------------------------------------------------------</td>
<td>---------------------------</td>
<td>------------------</td>
<td>--------------------------------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td><strong>Rheumatology:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLE (PRINTO)</td>
<td>Global assessment-physician; parent/child; activity tool; AntiDNA antibody; 24 hr proteinuria Serum creatinine</td>
<td>Global damage tool-SDI; physician; Height and weight Pubertal stage: Tanners; menses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DM (PRINTO)</td>
<td>Global assessment-physician; parent/child; activity tool; Muscle strength; Laboratory: Muscle enzymes</td>
<td>Global assessment-physician; MDI Height and weight; Pubertal stage: Tanners; menses Muscle strength-CMAS</td>
<td>CHAQ</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Rheumatology:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Juvenile arthritis (Giannini)</td>
<td>Physician global assessment; Number of active joints; ESR</td>
<td>Number of joints with limited range of movement</td>
<td>Functional ability</td>
<td>Overall well being (Parent/child reported)</td>
<td></td>
</tr>
</tbody>
</table>
Appendix 5 – Search strategy used in the systematic review, described in Chapter 3, of RCTs of inhaled corticosteroids for children with asthma

CENTRAL was searched in January 2008 (no language restrictions) using the following search strategy:

#1 asthma:ti,ab,kw
#2 (antiasthma OR anti-asthma):ti,ab,kw
#3 wheez*:ti,ab,kw
#4 (bronch?spas* OR bronchoconstric* OR bronchismus OR bronchiospas*):ti,ab,kw
#5 cough:ti,ab,kw
#6 (#1 OR #2 OR #3 OR #4 OR #5)
#7 (child*):ti,ab,kw
#8 (paediatric* OR pediatric):ti,ab,kw
#9 (infan*):ti,ab,kw
#10 (young*):ti,ab,kw
#11 (toddler*):ti,ab,kw
#12 bab*:ti,ab,kw
#13 (preschool or pre-school):ti,ab,kw
#14 (teenage*):ti,ab,kw
#15 (adolesce*):ti,ab,kw
#16 (#7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15)
#17 beclomet?asone:ti,ab,kw
#18 betamet?asone:ti,ab,kw
#19 fluticasone:ti,ab,kw
#20 budesonide:ti,ab,kw
#21 (corticosteroid* OR *corticoid*):ti,ab,kw
#22 (inhaled *steroid*):ti,ab,kw
#23 (pulmicort or azmacort or becoride or flixotide or flovent or aerobid or aerobec or qvar or vanceril or triamciclone):ti,ab,kw
#24 (#17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23)
#25 (#6 AND #16 AND #24)
Appendix 6 – Phase 1 questionnaire, and invitation, sent to clinicians

Dear BPRS member

My name is Ian Sinha, Specialist Registrar in paediatrics in the Mersey Deanery. I am a clinical research fellow in the Medicines for Children Research Network Clinical Trials Unit, at Alder Hey Childrens’ NHS Foundation Trust, and I am undertaking a PhD. My thesis (supervised by Professor Rosalind Smyth and Professor Paula Williamson) relates to outcomes in children with asthma. I hope to identify which outcomes should be measured in randomised controlled trials (RCTs) of chronic therapies for children with asthma. Your help in this project would be invaluable and I am very grateful to Professor Warren Lenney and the BPRS Executive Committee for contacting you on my behalf.

How you can help

In this study I hope to identify which outcomes are most important to clinicians who have experience of the outpatient management of children with asthma. I would be grateful if you would spare a few minutes to complete a very brief questionnaire, which can be found on the BPRS website (www.bprs.co.uk/asthmaquestionnaire.doc). You will be asked to list up to 5 beneficial or harmful outcomes of treatment that you find clinically important in school aged children (aged 5 to 18) and preschool children (less than 5 years). These should be things that you consider, when deciding whether to continue a child on treatment or alter a child’s regular asthma regime. The deadline for completing this questionnaire is Friday 8 May 2009.

What happens then?

If you complete this questionnaire, and are happy to be contacted again, you will be invited to complete a second questionnaire. On
this, you will be asked to score the importance of the outcomes suggested in the first questionnaire. 

In the final study report, I would like to acknowledge, by name, participants who complete both questionnaires. For this reason only, you will be asked to provide your name, job title, and the institution in which you work. Your details will not be passed on to any other party. The other participants will not know the answers which you, personally, provide.

**Rationale for this study**

Consensus guidelines on outcomes for children with asthma do not exist. In a systematic review of asthma RCTs, I found marked variations between studies in outcomes measured and the ways in which they were measured. I hope this study will contribute to the standardisation of selection and measurement of outcomes in RCTs in children with asthma.

**Funding**

I am funded by the NIHR Medicines for Children Research Network Clinical Trials Unit and Co-ordinating Centre.

**Ethics**

NRES have provided the following statement: “Based on the information you provided, our advice is that the project is not considered to be research. Therefore it does not require ethical review by a NHS Research Ethics Committee. We would deem this a service evaluation and development.”

**Thank you**

Thank you for taking the time to read this. I am very grateful and hope you will consider taking part. If you have any questions please do not hesitate to contact me. (iansinha@liv.ac.uk). The questionnaire is now ready for completion, and can be found on the BPRS website (www.bprs.co.uk/asthmaquestionnaire.doc).
Asthma Questionnaire for Dr Ian Sinha

Please answer the following two questions:
When you see children with asthma in clinic, you make an assessment as to whether their treatment is working. Please list UP TO FIVE beneficial or harmful outcomes of treatment that you find clinically most important in **school aged children** (aged 5 to 18). These factors should be things that you consider, when deciding whether to recommend continuing on current treatment or altering a child’s regular asthma therapy regime.

Now please list UP TO FIVE beneficial or harmful outcomes of treatment that you find clinically most important in **pre-school aged children** (aged less than 5). These factors should be things that you may consider, when deciding whether to recommend continuing on current treatment or altering a child’s regular asthma therapy regime.

Thank you for your time. From your answers, and those from a group of children with asthma and their parents, we will compile a list of outcomes suggested by a pre-defined minimum number of participants. This will form the basis for a second questionnaire, on which participants will be asked to score the importance of each outcome.

Would you like the study facilitator (Ian Sinha, SpR in paediatrics) to contact you by email when this second questionnaire is ready?  
Yes [ ] No [ ]

If you answered YES, please provide your email address here.

Your email address will NOT be passed on to ANY other party under any circumstances. We will not use it to contact you about any other studies. You would receive one email to confirm that your questionnaire was received, and one to inform you that the second
questionnaire is ready for completion. After that, we would only contact you once, to inform you when the final study report is ready for publication. After this, your email address would be permanently deleted from our records.

For acknowledgement purposes, we would like to know your name, job title and institution. These will not be passed to anyone else, and will be kept completely confidential, until the final study report is ready for publication. If you do not want to be acknowledged in this way, please leave the name and institution spaces blank, but for data analysis purposes, please let us know your job title.

<table>
<thead>
<tr>
<th>NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>JOB TITLE</td>
</tr>
<tr>
<td>INSTITUTION</td>
</tr>
</tbody>
</table>

PLEASE SEND YOUR COMPLETED QUESTIONNAIRE TO IAN SINHA BY EMAIL:
Appendix 7 - Phase 1 questionnaire distributed to parents

This questionnaire is about the daily preventer treatment your child takes to control their asthma.

In order to know whether medications for young people with asthma are helpful and safe, they must be tested by researchers. These researchers measure whether medications affect children in a number of ways. For example, they could test the following effects of a medication:

1. Whether it improves how well a child’s asthma is controlled in the short term, eg by reducing coughing or wheezing
2. Whether it improves the ability of children to do normal activities, eg playing sports or going to school
3. Whether it prevents long term problems associated with having asthma as a child, eg by preventing a child having asthma when they are an adult
4. Whether it reduces the number of times children are admitted to hospital with asthma
5. Whether it generally improves life for children and their families
6. Whether it has side effects.

We believe the good and bad effects of medications that researchers measure should be things that are important to children with asthma, and their parents. We would be grateful if you could complete this quick questionnaire. If you do, your answers will be kept completely confidential.

**Question 1**
How old is your child? _______________

**Question 2**
*Over the last twelve months*, have you generally felt that the regular preventer treatment that your child takes has kept their asthma under control?

Tick ONE box: Yes [ _ ]No [ ]

If you ticked YES, please tell us what aspects of your child’s asthma, or their daily life, have made you feel happy that they are on the correct regular medication.

If you ticked NO, please leave this question blank

____________________________________________________
____________________________________________________
____________________________________________________
____________________________________________________
____________________________________________________
Question 3

Over the last twelve months, have there been times when you felt that your child’s regular preventer treatment should be increased or changed, because their asthma was not under control?

**Tick ONE box: Yes [ ] No [ ]**

If you ticked YES, please tell us the reasons why you were not satisfied with the regular preventer treatment that they were taking?

If you ticked NO, please leave this space blank

____________________________________________________

____________________________________________________

____________________________________________________

____________________________________________________

Question 4

Does anything worry you about the fact that your child has asthma?

**Tick ONE box: Yes [ ] No [ ]**

If you ticked YES, please tell us the worries you have about the fact your child has asthma

If you ticked NO, please leave this space blank

____________________________________________________

____________________________________________________

____________________________________________________

____________________________________________________

Question 5

Does anything worry you about the regular preventer treatment that your child takes for their asthma?
Tick ONE box: Yes [ ] No [ ]

If you ticked YES, please tell us what worries you have about the treatment your child takes for their asthma. Please be as specific as you can.

If you ticked NO, please leave this question blank

______________________________________________________
______________________________________________________
______________________________________________________
______________________________________________________
Appendix 8 – Questionnaires distributed to clinicians and parents in Phase 2
Thank you for participating in the second part of our study

Before completing the questionnaire, please read this important background information:

Clinical trials are research studies. They are conducted to determine which treatments are best for patients with a given condition.

In clinical trials we aim to measure various effects of treatments, both beneficial and harmful, over a period of time. We refer to this as measuring the effects of treatments on various outcomes.

We would like to know which outcomes you feel are important, and should be measured in clinical trials of regular preventer treatments for children with asthma.

We will be asking you to consider outcomes in relation to pre-school and school age children separately.

Regular treatments for children can have a variety of beneficial effects, each of which could be measured as an outcome in clinical trials.

Please score how important each of the following outcomes are on a scale of 0-4 in relation to SCHOOL-AGE CHILDREN (5 years or older)

0= not at all important
(you are not concerned if this is measured in clinical trials or not)

4= very important
(you feel this is important and should be measured in clinical trials)

1. The effect of a treatment on daytime asthma symptoms
   (eg cough, wheeze, shortness of breath)
   0 1 2 3 4

2. The effect of a treatment on night-time asthma symptoms
   (eg coughing during sleep)
   0 1 2 3 4

3. The effect of a treatment on activity or exercise without asthma symptoms
   0 1 2 3 4

4. The effect of a treatment on how often children need to use their reliever inhaler
   0 1 2 3 4

5. The effect of a treatment on how often children have exacerbations of asthma that require a course of oral steroids (prednisolone)
   0 1 2 3 4

6. The effect of a treatment on a child’s overall asthma control (as judged by the parents and the child, rather than by doctors)
   0 1 2 3 4

Please tell us your FULL NAME*
7. The effect of a treatment on lung function tests
   0 0 0 1 0 2 0 3 0 4

8. The effect of a treatment on preventing asthma-related death
   0 0 0 1 0 2 0 3 0 4

9. The effect of a treatment on how well children can do normal activities that they would like to do (e.g., playing sport, or socialising with friends)
   0 0 0 1 0 2 0 3 0 4

10. The effect of a treatment on the amount of school a child misses because of asthma
   0 0 0 1 0 2 0 3 0 4

11. The effect of a treatment on the overall quality of life of children with asthma
    0 0 0 1 0 2 0 3 0 4

12. The effect of a treatment on how frequently children visit a GP or A&E department because of asthma
    0 0 0 1 0 2 0 3 0 4

13. The effect of a treatment on how frequently children need to be admitted to hospital because of asthma
    0 0 0 1 0 2 0 3 0 4

14. The effect of a treatment on health-related problems when children are older
    0 0 0 1 0 2 0 3 0 4

Please score how important you think the following side effects are for SCHOOL AGE CHILDREN

15. Short term side effects that stop when the treatment stops
    0 0 0 1 0 2 0 3 0 4

16. Effects of treatment on growth
    0 0 0 1 0 2 0 3 0 4

17. Other long-term side effects that continue, or appear, after the treatment stops
    0 0 0 1 0 2 0 3 0 4

18. You have just scored the outcomes listed below.

Please tick the THREE MOST IMPORTANT in relation to SCHOOL-AGE children:

- The effect of a treatment on daytime symptoms
- The effect of a treatment on night-time symptoms
- The effect of a treatment on activity or exercise without symptoms
- The effect of a treatment on how often children need to use their reliever inhaler
- The effect of a treatment on exacerbations of asthma that require oral steroids
- The effect of a treatment on overall asthma control (judged by parent and child)
- The effect of a treatment on lung function tests
- The effect of a treatment on preventing asthma-related death
- The effect of a treatment on how well a child can do normal activities
- The effect of a treatment on school missed because of asthma
- The effect of a treatment on a child's overall quality of life
- The effect of a treatment on visits to a GP or A&E department
- The effect of a treatment on admissions to hospital
- The effect of a treatment on health-related problems when a child is older
- Short term side effects when the treatment stops
- Effects of treatment on growth
- Other long-term side effects that continue or appear after the treatment stops

19. Are there any other outcomes of treatment which are not listed that would have made it into your top 3?
Thank you

Now please score how important each of the following outcomes are on a scale of 0-4 in relation to PRE-SCHOOL CHILDREN (younger than 5 years)

0= not at all important (you are not concerned if this is measured in clinical trials or not)
4= very important (you feel this is important and should be measured in clinical trials)

20. The effect of a treatment on daytime asthma symptoms (eg cough, wheeze, shortness of breath)

21. The effect of a treatment on night-time asthma symptoms (eg coughing during sleep)

22. The effect of a treatment on activity or exercise without asthma symptoms

23. The effect of a treatment on how often children need to use their reliever inhaler

24. The effect of a treatment on how often children have exacerbations of asthma that require a course of oral steroids (prednisolone)

25. The effect of a treatment on a child's overall asthma control (as judged by the parents and the child, rather than by doctors)

26. The effect of a treatment on preventing asthma-related death

27. The effect of a treatment on how well children can do normal activities that they would like to do (eg playing sport, or socialising with friends)

28. The effect of a treatment on the amount of school/nursery a child misses because of asthma

29. The effect of a treatment on the overall quality of life of children with asthma

30. The effect of a treatment on the impact of asthma on the whole family (eg number of days of work that parents miss because of a child's asthma, or parental anxiety)

31. The effect of a treatment on how frequently children visit a GP or A&E department because of asthma

32. The effect of a treatment on how frequently children need to be admitted to hospital because of asthma

33. The effect of a treatment on health-related problems when children are older

Please score how important you think the following side effects are for PRE-SCHOOL CHILDREN

34. Short term side effects that stop when the treatment stops

35. Effects of treatment on growth

36. Other long-term side effects that continue, or appear, after the treatment stops
37. You have just scored the outcomes listed below.

Please tick the THREE MOST IMPORTANT in relation to PRE-SCHOOL children* 

☐ The effect of a treatment on daytime symptoms
☐ The effect of a treatment on night-time symptoms
☐ The effect of a treatment on activity or exercise without symptoms
☐ The effect of a treatment on how often children need to use their reliever inhaler
☐ The effect of a treatment on exacerbations of asthma that require oral steroids
☐ The effect of a treatment on overall asthma control (judged by parent and child)
☐ The effect of a treatment on preventing asthma-related death
☐ The effect of treatment on how well a child can do normal activities
☐ The effect of a treatment on school/nursery missed because of asthma
☐ The effect of a treatment on overall quality of life
☐ The effect of a treatment on the overall impact of asthma on the family
☐ The effect of a treatment on visits to a GP or A+E department
☐ The effect of a treatment on admissions to hospital
☐ The effect of a treatment on health-related problems when a child is older
☐ Short term side effects when the treatment stops
☐ Effects of treatment on growth
☐ Other long-term side effects that continue or appear after the treatment stops

38. Are there any other outcomes of treatment which are not listed that would have made it into your top 3?


Dear Parent,

We are very grateful that you have agreed to participate in our study.

Before completing the questionnaire, please read this important background information:

Clinical trials are research studies. They are conducted to determine which treatments are best for patients with a given condition.

In clinical trials we aim to measure various effects of treatments, both beneficial and harmful, over a period of time. We refer to this as measuring the effects of treatments on various outcomes.

We would like to know which outcomes you, as a parent, feel are important, and should be measured in clinical trials of regular preventer treatments for children with asthma.

How old is your child, in years?*

Please score how important each of the following outcomes are on a scale of 0-4

0= not at all important (you are not concerned if this is measured in clinical trials or not)

4= very important (you feel this is important and should be measured in clinical trials)

1. The effect of a treatment on daytime asthma symptoms
   (eg cough, wheeze, shortness of breath)
   ○ 0 1 2 3 4

2. The effect of a treatment on night-time asthma symptoms
   (eg coughing during sleep)
   ○ 0 1 2 3 4

3. The effect of a treatment on activity or exercise without asthma symptoms
   ○ 0 1 2 3 4
4. The effect of a treatment on how often children need to use their reliever inhaler
   ○ 0 1 2 3 4

5. The effect of a treatment on how often children have exacerbations of asthma that require a course of oral steroids (prednisolone)
   ○ 0 1 2 3 4

6. The effect of a treatment on a child’s overall asthma control (as judged by the parents and the child, rather than by doctors)
   ○ 0 1 2 3 4

7. The effect of a treatment on lung function tests (blowing tests)
   [PLEASE SCORE THIS OUTCOME ONLY IF YOUR CHILD IS AGED FIVE YEARS OR OLDER]
   ○ 0 1 2 3 4

8. The effect of a treatment on preventing asthma-related death
   ○ 0 1 2 3 4

9. The effect of a treatment on how well children can do normal activities that they would like to do (eg playing sport, or socialising with friends)
   ○ 0 1 2 3 4

10. The effect of a treatment on the amount of school/nursery a child misses because of asthma
    ○ 0 1 2 3 4

11. The effect of a treatment on the overall quality of life of children with asthma
    ○ 0 1 2 3 4

12. The effect of a treatment on the impact of asthma on the whole family (eg number of days of work parents miss because their child’s asthma is bad)
    [PLEASE SCORE THE OUTCOME ‘IMPACT OF ASTHMA ON THE WHOLE FAMILY’ ONLY IF YOUR CHILD IS Younger THAN FIVE YEARS OF AGE]
    ○ 0 1 2 3 4

13. The effect of a treatment on how frequently children visit a GP or A&E department because of asthma
    ○ 0 1 2 3 4

14. The effect of a treatment on how frequently children need to be admitted to hospital because of asthma
    ○ 0 1 2 3 4

15. The effect of a treatment on health-related problems when children are older
    ○ 0 1 2 3 4
15. Treatments for children with asthma can also have side-effects.

16. **Short term side effects that stop when the treatment stops**
   - 0 0 1 2 3 4

17. **Effects of treatment on growth**
   - 0 0 1 2 3 4

18. **Other long-term side effects that continue, or appear, after the treatment stops**
   - 0 0 1 2 3 4

19. You have just scored the outcomes listed below.
   
   Please tick the THREE MOST IMPORTANT outcomes *
   
   - The effect of a treatment on daytime symptoms
   - The effect of a treatment on night-time symptoms
   - The effect of a treatment on activity or exercise without symptoms
   - The effect of a treatment on how often children need to use their reliever inhaler
   - The effect of a treatment on exacerbations of asthma that require oral steroids
   - The effect of a treatment on overall asthma control (judged by parent and child)
   - The effect of treatment on lung function (blowing) tests
   - The effect of a treatment on preventing asthma-related death
   - The effect of treatment on how well a child can do normal activities
   - The effect of a treatment on school/nursery missed because of asthma
   - The effect of a treatment on overall quality of life
   - The effect of a treatment on the impact of asthma on the whole family
   - The effect of a treatment on visits to a GP or A&E department
   - The effect of a treatment on admissions to hospital
   - The effect of a treatment on health-related problems when a child is older
   - Short term side effects when the treatment stops
   - Effects of treatment on growth
   - Other long-term side effects that continue or appear after the treatment stops

20. Are there any other outcomes of treatment which are not listed that would have made it into your top 3? 

   

## Appendix 9 – Summary of results from Asthma UK

Outcomes for pre-school children as scored by Parents from Asthma UK (n=7)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Median (IQR)</th>
<th>Number (%) of times outcome scored in Top 3</th>
<th>Number (%) of times outcome scored as 4/4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daytime symptoms</td>
<td>4 (3,4)</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Nocturnal symptoms</td>
<td>4 (4,4)</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Activity or exercise</td>
<td>3 (2,4)</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Use of reliever inhaler</td>
<td>3 (3,4)</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Exacerbations</td>
<td>4 (3,4)</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Parent/child overall assessment</td>
<td>4 (3,4)</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Death</td>
<td>4 (3,4)</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Normal activities</td>
<td>4 (3,4)</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>School</td>
<td>4 (3,4)</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Quality of Life</td>
<td>4 (3,4)</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Family outcomes</td>
<td>4 (3,4)</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>GP/A+E attendance</td>
<td>4 (3,4)</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Hospital admission</td>
<td>4 (4,4)</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Health related problems when older</td>
<td>4 (3,4)</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Short term adverse effects</td>
<td>4 (3,4)</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Growth</td>
<td>4 (3,4)</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Long term adverse effects</td>
<td>4 (3,4)</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

Outcomes for school-aged children as scored by Parents from Asthma UK (n=6)

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Median (IQR)</th>
<th>Number (%) of times outcome scored in Top 3</th>
<th>Number (%) of times outcome scored as 4/4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daytime symptoms</td>
<td>4 (3,4)</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Nocturnal symptoms</td>
<td>4 (3,4)</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Activity or exercise</td>
<td>2.5 (1,3)</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Use of reliever inhaler</td>
<td>4 (3,4)</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Exacerbations</td>
<td>4 (4,4)</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Parent/child overall assessment</td>
<td>4 (4,4)</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Lung function tests</td>
<td>3 (3,4)</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Death</td>
<td>4 (4,4)</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Normal activities</td>
<td>4 (3,4)</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>School</td>
<td>3 (3,4)</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Quality of Life</td>
<td>4 (4,4)</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>GP/A+E attendance</td>
<td>3 (3,3)</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Hospital admission</td>
<td>4 (3,3)</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Health related problems when older</td>
<td>4 (3,3)</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Short term adverse effects</td>
<td>3 (2,4)</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Growth</td>
<td>4 (3,4)</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Long term adverse effects</td>
<td>4 (4,4)</td>
<td>1</td>
<td>4</td>
</tr>
</tbody>
</table>
Appendix 10 – Correspondence with National Research Ethics Service

From: Sinha, Ian [mailto:I.Sinha@liverpool.ac.uk]
Sent: 26 November 2008 15:23
To: NRES Queries Line
Cc: Williamson, Paula; Smyth, Rosalind

Subject: Advice regarding ethical review for survey of asthma outcomes

Dear COREC

I am a clinical research fellow at the Medicines for Children Research Network Clinical Trials Unit, Liverpool. We are currently planning a piece of work to standardise, and improve, the outcomes which are measured in clinical trials of children with asthma. The work will be entitled “Development of a core set of outcomes for use in clinical trials in children with asthma”. Ultimately we aim to conduct a Delphi survey of clinicians who are involved in clinical research, to come to a consensus regarding what should be measured in clinical trials in asthma. In the first phase of our study we wish to survey parents of children with asthma, and children themselves (where appropriate), regarding their views on which outcomes should be measured in clinical trials. An outline of this project is listed below:

Format: Anonymous questionnaire
Participants:
(a) Parents of children attending asthma clinics at Alder Hey Childrens’ NHS Foundation Trust
(b) Children, aged at least 10 years old, attending asthma clinics at Alder Hey Childrens’ NHS Foundation Trust
(c) Parents of children with asthma who are involved in the Asthma UK research and policy groups. We will not be contacting children through Asthma UK.

Confidentiality and anonymity of all participants will be maintained.
We need to consider the potential timeframe of this study, and therefore we need to know whether or not we require ethical review. Having seen the COREC ethics e-group guideline “Differentiating audit, service evaluation and research” (November 2006), it would seem that our survey is not designated as research. We would, however, appreciate your advice regarding our interpretation of the guidance, with particular reference to the following question:

As this is an anonymised survey of attitudes towards clinical trials, we would anticipate that ethical review would not be required. Is this correct?
I look forward to hearing from you
Kind regards
Yours sincerely
Ian Sinha

RE: Advice regarding ethical review for survey of asthma outcomes
NRES Queries Line [queries@nres.npsa.nhs.uk]

Sent: 28 November 2008 10:56
To: Sinha, Ian
Your query was reviewed by our Queries Line Advisers.
Our leaflet “Defining Research”, which explains how we differentiate research from other activities, is published at:
http://www.nres.npsa.nhs.uk/rec-community/guidance/#researchoraudit

Based on the information you provided, our advice is that the project is not considered to be research according to this guidance. Therefore it does not require ethical review by a NHS Research Ethics Committee. I would deem this a service evaluation and development
If you are undertaking the project within the NHS, you should check with the relevant NHS care organisation(s) what other review arrangements or sources of advice apply to projects of this type. Guidance may be available from the clinical governance office.
Although ethical review by a NHS REC is not necessary in this case, all types of study involving human participants should be conducted in accordance with basic ethical principles such as informed consent and respect for the confidentiality of participants. When processing identifiable data there are also legal requirements under the Data Protection Act 2000.
When undertaking an audit or service/therapy evaluation, the investigator and his/her team are responsible for considering the ethics of their project with advice from within their organisation. University projects may require approval by the university ethics committee.
This response should not be interpreted as giving a form of ethical approval or any endorsement of the project, but it may be provided to a journal or other body as evidence that ethical approval is not required under NHS research governance arrangements.
However, if you, your sponsor/funder or any NHS organisation feel that the project should be managed as research and/or that ethical review by a NHS REC is essential, please write setting out your reasons and we will be pleased to consider further.
Where NHS organisations have clarified that a project is not to be managed as research, the Research Governance Framework states that it should not be presented as research within the NHS.
Regards
Queries Line
National Research Ethics Service
National Patient Safety Agency
4-8 Maple Street, London, W1T 5HD, Ref: 04/02, **This reply may have been sourced in consultation with other members of the NRES team.
Appendix 11 – Copy of publications arising from the work in this thesis
A Systematic Review of Studies That Aim to Determine Which Outcomes to Measure in Clinical Trials in Children

Ian Sinha¹*, Leanne Jones¹, Rosalind L. Smyth¹, Paula R. Williamson²

¹ Institute of Child Health, University of Liverpool, Liverpool, United Kingdom, ² Centre for Medical Statistics and Health Evaluation, University of Liverpool, Liverpool, United Kingdom

ABSTRACT

Background

In clinical trials the selection of appropriate outcomes is crucial to the assessment of whether one intervention is better than another. Selection of inappropriate outcomes can compromise the utility of a trial. However, the process of selecting the most suitable outcomes to include can be complex. Our aim was to systematically review studies that address the process of selecting outcomes or outcome domains to measure in clinical trials in children.

Methods and Findings

We searched Cochrane databases (no date restrictions) in December 2006; and MEDLINE (1950 to 2006), CINAHL (1982 to 2006), and SCOPUS (1966 to 2006) in January 2007 for studies of the selection of outcomes for use in clinical trials in children. We also asked a group of experts in paediatric clinical research to refer us to any other relevant studies. From these articles we extracted data on the clinical condition of interest, description of the method used to determine which outcomes to measure in clinical trials in children.

Conclusions

Very few studies address the appropriate choice of outcomes for clinical research with children, and in most paediatric specialties no research has been undertaken. Among the studies we did assess, very few involved parents or children in selecting outcomes that should be measured, and none directly involved children. Research should be undertaken to identify the best way to involve parents and children in assessing which outcomes should be measured in clinical trials.

* To whom correspondence should be addressed. E-mail: i.sinha@liv.ac.uk

The Editors’ Summary of this article follows the references.
Introduction

The purpose of a clinical trial is to determine the benefits and harms of an intervention. This determination is made by measuring the effects of different treatments on outcomes. The selection of appropriate outcomes, therefore, is crucial to the assessment of whether one intervention is better than another. This review relates to studies that explain how outcomes have been selected for use in clinical trials in children younger than 16 years of age. For the purposes of this review we define children by age rather than by the literal meaning of offspring.

What Outcomes Measure—The Impact of Illness on a Patient’s Life

Models have been developed that describe the effects of a disease on a patient, for example the biopsychosocial model and the World Health Organisation framework of impairment, disability, and handicap [1,2]. Although these models differ in many ways, an underlying theme is that illnesses affect more than one aspect of a patient’s life. For example, asthma may affect a child’s life by way of troublesome daily symptoms even when the child is “well,” exacerbations, disrupted school attendance, and abnormal lung function. Each of these effects of asthma on the child’s life is potentially amenable to improvement after starting an intervention. In clinical trials, the extent to which an intervention affects the impact of an illness on a patient’s life is reflected by measuring change in outcomes.

For the purpose of this review, we clarify our terminology in Table 1.

Outcomes can reflect various effects of an intervention. They may directly measure a definitive clinical change, such as death or hospital admission. Surrogate outcomes, which are sometimes used in lieu of a definitive clinical outcome, are an amalgamation of more than one outcome within an outcome domain, such as a score based on a variety of symptoms. More than one outcome measure may be possible to use to represent change in an outcome. To wait for the clinical change to actually occur. In other words, they are proximal to the clinical outcome on the disease pathway, so a change can be detected sooner. They may be a measure of intermediate health status, which may be used to predict future health status; for example, glycosylated haemoglobin is used as a measure of current disease control in patients with diabetes mellitus, and has been shown to be a useful predictor of future control [3]. A surrogate outcome may even be an assumed or established risk factor that actually impacts on disease progression; for example, neonatal intraventricular haemorrhage, which is a recognised complication of prematurity, is thought to alter brain development in the early stages of life and predispose babies to developmental problems in childhood. There are validation criteria that should be fulfilled before a surrogate outcome can be confidently used in place of a definitive clinical outcome in clinical trials [4].

An outcome domain may be represented by a variety of outcomes. The domain of health care utilisation, for example, may be reflected by number of visits to a general practitioner, number of hospital admissions, or days spent in hospital. Conversely, outcomes may be relevant to more than one domain. For example, in clinical trials of children with asthma, the outcome “number of courses of rescue prednisolone therapy” may be a measure of health care utilisation, or could alternatively represent change in the domain “exacerbations.” These various “levels” of outcome measurement are illustrated schematically in Figure 1.

Selecting Outcomes for Use in Clinical Trials

Clinical trials are “only as credible as their outcomes” [5], so when designing a clinical trial, the decision as to which outcomes should be measured is crucial. The selection of inappropriate outcomes can lead to wasted resources or misleading information that overestimates, underestimates, or completely misses the potential benefits of an interven-

---

**Table 1. Definitions Used in Review**

<table>
<thead>
<tr>
<th>Terminology</th>
<th>Description</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outcome domain</td>
<td>A relatively broad aspect of the effect of illness on a child, within which an improvement may occur in response to an intervention. In general these domains may not be directly measurable themselves, so outcomes are selected to assess change within them.</td>
<td>(1) In clinical trials of children with diabetes, outcome domains may include acute metabolic complications, long-term glucose control, renal damage, or effects on schooling. (2) In clinical trials of children with asthma, outcome domains may include lung function, health care utilisation, and symptom control</td>
</tr>
<tr>
<td>Outcome(endpoint)</td>
<td>A measurable variable within an outcome domain. The outcome can be measured at a variety of time points, which must be clearly stated by authors of clinical trials.</td>
<td>(1) Absolute FEV1 expressed as change from baseline (2) Number of admissions to hospital within a six month period (3) Time to first seizure after starting an antiepileptic intervention</td>
</tr>
<tr>
<td>Composite outcome</td>
<td>An outcome that encompasses more than one domain. A composite outcome may span more than one outcome domain. These are sometimes referred to as global outcomes, because they theoretically measure several aspects of the impact of illness on a patient.</td>
<td>In the NIMH Collaborative Multisite Multimodal Treatment Study of Children with Attention Deficit/Hyperactivity Disorder, response to treatment was measured by a composite outcome comprising outcomes that reflect clinical outcome, academic functioning, effects on family life, and other domains [37]</td>
</tr>
<tr>
<td>Outcome measure</td>
<td>A scale, scoring system, questionnaire, or other tool used for measuring an outcome. They may be an amalgamation of more than one outcome within an outcome domain, such as a score based on a variety of symptoms. More than one outcome measure may be possible to use to represent change in an outcome.</td>
<td>(1) The Paediatric Asthma Quality of Life Questionnaire [38] (2) Childhood Health Assessment Questionnaire [39] (3) Gastro-esophageal Reflux Disease Symptom Questionnaire [40]</td>
</tr>
</tbody>
</table>
tation. Examples of these problems are well documented [6,7]. Investigators can select from a range of several potential outcomes spanning different domains when they are designing a clinical trial, however, and the process of determining which outcome to use can be complex. The difficulty of selecting the most appropriate outcomes for use in a clinical trial is reflected in the fact that in several fields of clinical research there is much heterogeneity between clinical trials of specific diseases regarding exactly which outcomes to select [8,9]. Some of the factors underlying this uncertainty may be that for these conditions there is uncertainty about which outcome domains are most relevant to patients, that the performance characteristics of potential outcomes have not been established, or that as the general care of patients has improved, previously used outcomes are no longer relevant.

Since the late 1980s there have been attempts, notably in the field of rheumatology (Outcome Measures in Rheumatology, http://www.omeract.org/), to develop “core sets” of outcomes that should be measured in all clinical trials of specific conditions. These studies generally use techniques to ascertain a consensus opinion from clinical experts as to which outcomes are most suitable for use in clinical trials. The three commonly used consensus techniques are nominal group technique (NGT), which entails structured face-to-face discussion with the aim of developing a solution to a specific problem, followed by a vote on the issue; Delphi technique, in which opinions are sought from individuals and the collated results are fed back to the group as a whole, to generate further discussion and finally reach an agreement; and semistructured discussion based around broader discussion points.

The objective of this project was to systematically review studies that address the process of selecting which outcome domains or outcomes to measure in clinical trials in children under 16 years of age. We have restricted this review to studies that address the process of selecting which outcomes to use, in addition to the consensus techniques described above, may be to ascertain the opinions of both children and their parents regarding what they think are important aspects of their disease. This process poses unique challenges that may not be relevant when conducting similar research in adults with a particular condition, so it is appropriate to specifically review studies pertaining to outcome domain and outcome selection in clinical trials in children.

Methods

Study Selection

Included studies. We decided that the following types of studies would be eligible for inclusion in the review: (1) Studies that develop or apply methodology for selecting outcome domains or outcomes to be used in clinical trials in children younger than 16 years of age, and (2) systematic reviews of these articles.

Excluded studies. We excluded the following types of studies. (1) Studies that do not specifically state that the outcomes are appropriate for use in a paediatric population. (2) Studies that discuss how to measure, rather than how to select, an outcome domain or outcome for use in clinical trials. This category includes studies discussing performance characteristics of outcomes or instruments for measuring them. (3) Studies relating to clinical trials that assess interventions given to adults by measuring outcomes in children, for example the selection of neonatal outcomes to assess care given to their mothers.

Identification of Studies

In December 2006 we searched Cochrane databases (no date restrictions), and in January 2007 we searched MEDLINE (1950 to 2006; http://www.ovid.com/site/catalog/DataBase/901.jsp?top=2&mid=3&bottom=7&subsection=10), CINAHL (1982 to 2006; http://www.cinahal.com/), and SCOPUS (1966 to 2006; http://www.scopus.com/). SCOPUS is a platform that enables the searching of several databases, including EMBASE, simultaneously. We used the following abbreviated search strategy: “Clinical trials” AND “Outcomes” AND “Children” AND “methodology”. Details of the full search strategy are included in Table S1.

The abstracts produced by the searches were initially screened twice by one reviewer. The full texts of all potentially relevant articles were obtained, and these were assessed with regard to the eligibility criteria. Data were extracted from the studies that met all the eligibility criteria.

A second reviewer, who was blinded to the first reviewer’s assessment of the abstracts, independently screened a database of abstracts that comprised all the abstracts for which the first reviewer obtained full text, plus a selection of abstracts rejected at the initial screening stage by the first reviewer. The purpose of this approach was to check the sensitivity of the initial screening process that had been performed in full by the first reviewer. A sample, rather than the complete set, was selected due to resource constraints. Any disagreements between the reviewers were resolved by discussion.

This process led to a list of studies for which full text were obtained. Both reviewers then scrutinised these articles for the predetermined inclusion and exclusion criteria in order to determine which studies should be included in the systematic review.

We then emailed a list of the studies we had identified to the Clinical Study Group (CSG) members of the Medicines for Children Research Network (MCRN) and asked if they knew of any other relevant studies, published or unpublished, that should be included. The CSG constitutes a multidisciplinary group of clinical experts with a strong interest in the planning of clinical trials within their specialities.

Data Extraction

The following data were extracted by one reviewer (IS) and checked independently by the second reviewer (LJ): (1)
Condition for which the outcome domains or outcomes are discussed; (2) Description of the method; (3) People involved in selecting outcomes or domains; (4) Outcome domains or outcomes selected; (5) The geographical setting of the collaborations, ascertained either by reading the text or, where listed, the names and institutions of people involved in the collaborations; (6) Limitations of the method as defined by the authors.

Assessment of Methodological Quality

The methodological quality of the studies was assessed by one author (IS). If a study developed or used methodology to select an outcome domain or an outcome, the article was assessed in terms of whether the method was described in sufficient detail to allow a reader to utilise it.

If a study described a consensus procedure, the following points were noted: (1) Is the selection process and areas of expertise of the participants described? (2) Is the process of coming to consensus described in detail?

We searched for a validated assessment tool for critically appraising consensus statements but we could not identify one. We therefore asked two experts, one with experience of qualitative research and the other with experience of participating in a consensus statement exercise to advise on this methodological assessment checklist.

For systematic reviews of studies which used methodology for selecting outcomes it was agreed that we would use the Critical Appraisal Skills Programme Systematic Review Appraisal tool for assessing their methodological quality (http://www.phru.nhs.uk/Pages/PHD/resources.htm).

Data Analysis and Presentation of Results

For synthesis of data we described the studies narratively and tabulated their characteristics. Consistent with the nature of the data, the results are presented in textual format.

Results

Description of Studies

The initial database search identified 8,889 potentially relevant abstracts, of which 70 articles were retrieved in full and, finally, 25 included in the full review, as depicted by the flowchart in Figure 2.

In total, 57 full-text articles were reviewed and subsequently excluded. Of the 57 studies 19 were excluded because the authors did not use methodology for selecting outcomes (e.g., a review article based on personal opinion), 18 because the study related to how to measure outcomes rather than which ones to select, ten because the study made no mention of outcome selection, six because the study did not specifically state that the outcomes which were selected were relevant to children, and four that described consensus statements relating to clinical practice rather than clinical trial design. The reasons for exclusion of each individual study are presented in Table S2.

In addition, 13 specific articles were suggested by the members of the MCRN CSGs in response to our email query. One of these articles summarised the work of a collaboration that had been identified by the literature search but did not describe the methodology used by the group in sufficient detail to warrant inclusion in the full review, so it is added as an additional reference [10]. The other studies identified were not deemed to be eligible for the full review.

Agreement between Reviewers

The second reviewer was provided with a database of 100 abstracts. These included 70 for which the first reviewer thought full text should be retrieved, and a randomly selected sample of 30 abstracts that had been excluded by the first reviewer at the abstract screening stage.

The second reviewer agreed that all 30 abstracts rejected at the abstract screening stage were appropriately excluded by the first reviewer.

Of the 70 abstracts for which full text was obtained by the first reviewer, the second reviewer agreed with 61, and disagreed with nine. After discussion it was agreed that all nine should be retrieved in full based on the abstract. Of these, eight were excluded after reading the full text and one was included.

Following full text review there was complete agreement between the second and first reviewer about the 25 included and 45 excluded abstracts. The second reviewer also checked all the data that had been extracted by the first reviewer, and agreed completely with the tabulated characteristics of the studies.

Included Studies

The 25 articles included in the full review represented the work of 13 collaborative groups. The characteristics of each study are included in Table S3, and summarised in Table 2.

Six of these groups (Griffiths et al. [11], Ramsey et al. [12], Pavletic et al. [13], Giannini et al. [14], International Myositis Assessment and Clinical Studies group (IMACS) [15], and Paediatric Rheumatology International Trials Organisation (PRINTO) [16]) aimed to develop a consensus statement specifically about outcome measures that should be used in clinical trials of certain medical conditions. Five groups (Carlson et al. [17], Goldstein et al. [18], LaFrance et al. [19], Neonatal Drug Development Initiative (NDDI) [20–24], and West Delphi group [25]) discussed which outcomes to measure as part of workshops which addressed wider clinical trial design issues. One group (Smith et al. [26]) aimed to ascertain the opinions of clinical experts about which...
outcomes to measure in clinical trials in children with asthma. One group (DeRouen et al. [27]) ascertained the opinions of experts about which outcome to measure in a specific safety trial of two interventions used in paediatric dental restoration. Our search identified no systematic reviews of studies that had selected outcome measures for use in clinical trials.

Most groups appeared to comprise an international collaboration of participants. Eight groups were based in the US (Ramsey et al. [12], Goldstein et al. [18], La France et al. [19], the NDDI [20–24], Carlson et al. [17], Griffiths et al. [11], DeRouen et al. [27], and Pavletic et al. [13]). One group was based in Europe (West Delphi group [25]). One group was based in Australasia (Smith et al. [26]). The three rheumatology collaborations [14–16] seem to have been based mainly in the US, but it appears that many of the leaders of these groups were based in Europe.

Methodological Quality of Studies

General observations regarding the methodological quality of the studies are provided in this section. Methodological features of each specific study are provided in Table S4.

Reporting of methodology. Of the 13 collaborations four used structured techniques to formulate a consensus (Giannini et al. [14], West Delphi group [25], PRINTO [16], and IMACS [15]); these were NGT and/or Delphi technique. Of these groups, three described the process very clearly. Eight collaborations came to a consensus by structured discussion, but without using structured consensus formulation techniques mentioned above (Goldstein et al. [18], Ramsey et al. [12], LaFrance et al. [19], NDDI [20–24], Carlson et al. [17], Griffiths et al. [11], DeRouen et al. [27], and Pavletic et al. [13]). All of these groups described the discussions in some detail. One group sought opinions in a questionnaire-based survey, and the methodology used for this study was described in sufficient detail to be able to repeat the study (Smith et al. [26]).

Selection of participants. All groups described the background of their participants. Only two of these groups described in detail the process by which it was decided specifically which individuals would be involved (West Delphi group [25] and Smith et al. [26]).

Methods Used to Select Outcomes

The following techniques were used to ascertain expert opinion concerning which outcomes ought to be measured in clinical trials of children with specific conditions.

Delphi technique. As described earlier, Delphi technique is one method of reaching a consensus opinion that relies on one person collating the views of each individual in a group, collating the results, and feeding these back to the whole group [28]. Statements made by participants at each stage of the process can be used to formulate the next round of questions. This technique has been used since the 1950s. Three groups utilised this method as follows.

The West Delphi group [25] used this technique to develop a core set of outcomes for use in clinical trials of children suffering from infantile spasms. The whole process was conducted by email over six rounds. In round one a group of 133 invited participants, of which 42 responded, were asked multiple-choice questions covering various aspects of clinical trial design, including outcomes. In round two a separate set of multiple-choice questions was provided, having fed the results of round one back to the group. At this stage the participants were also invited to comment and provide their personal opinions regarding outcomes. In round three statements were formulated from those responses in rounds one and two that had represented majority opinion. Participants were invited to respond as to whether they agreed or disagreed with these statements. For round four the statements were modified, and participants com-
mented on their suitability and content. Rounds five and six consisted of formulation of a draft and, subsequently, a final paper that were altered following comments from the group.

The IMACS [15] group used a Delphi technique to develop a core set of outcome domains and outcomes for use in clinical trials in children with inflammatory myopathy. The actual process itself is not described in detail in the article, but authors stated that the group consisted of “more than 100” members.

The PRINTO [16] group used a Delphi technique over two sequential questionnaire-based surveys to identify which variables should be measured in clinical trials of children with SLE. In the first questionnaire they asked 267 participants to indicate up to ten variables they judged as clinically most important. In the second questionnaire, the facilitators listed those indicators that had been suggested by at least ten responders, and asked the participants to rank in order their top ten choices.

**Nominal group technique.** NGT is a technique based on structured face-to-face discussion developed in the early 1970s. Having discussed a problem, with a view to providing potential solutions, the participants vote on the options presented, and ultimately a consensus is reached [29]. Two groups utilised this technique.

PRINTO [16] used NGT to discuss specific issues regarding the potential outcomes identified by the initial Delphi technique discussions described earlier. The NGT exercise had five objectives, which were tackled by a group of 40 participants: (1) to classify the proposed outcomes into “domains”; (2) to classify the outcomes into “concepts of disease activity”; (3) to select the outcome domains that should be measured in clinical trials; (4) to select the outcomes that should be used to measure these domains; and (5) to discuss specific design issues of the prospective validation phase of the study.

Gianinni et al. [14] used NGT to select from a set of potential outcomes a preliminary core set of six. The process used is not described in further detail in the study. The initial list of potential outcomes had been identified by sending a questionnaire to a 16-member advisory council.

**Semistructured discussion.** Most groups did not use structured techniques of consensus development such as Delphi or NGT, but rather came to consensus by discussion at meetings or workshops. As mentioned earlier, some collaborations—for example those groups discussing methodology issues in studies of neonates—discussed outcome selection broadly, as part of wider discussions about neonatal clinical trial designs. Other groups—for example, the group selecting outcome measures for use in an individual clinical trial of dental restoration—conducted very focussed discussions about very specific problems.

**Questionnaires.** Smith et al. [26] sent questionnaires to 39 health care professionals and researchers with expertise in asthma to ask which outcomes they would use for a variety of clinical, research, and public health scenarios, including questions about which outcome they would use in clinical trials of acute and preventative asthma medication. Three groups (Gianinni et al. [14], West Delphi [25], and PRINTO [16]) used questionnaires as part of the process of ascertaining the opinions of experts, mainly in the preliminary phases of the consensus process.

**People Involved in Selecting Outcomes**

**Clinical experts.** All 13 groups included people with clinical expertise in the fields for which they were selecting outcomes. Eight groups specifically mention the involvement of clinicians in both pediatric and adult health care.

**Research experts.** All groups appeared to include members who were experienced in research in the clinical condition for which outcomes were being selected. In addition to these clinical research experts, some groups also included biostatisticians and epidemiologists. Three groups involved experts from other clinical research areas who had experience in collaborations that had selected outcomes for clinical trials of other medical conditions. More collaborations may have used experts from this category, but may have referred to them generically as “research experts,” so it is difficult to quantify exactly how many groups used this approach.

**Patients or parents.** Three groups ascertained the opinions of parents of children with medical conditions as to which outcomes they thought should be measured, but no group involved children directly. IMACS [15] involved two patient support group leaders who had a child who suffered from inflammatory myopathy. Although this was not explicitly stated in the text, we elicited this information by searching for the names of the support group leaders on an internet search engine. Carlson et al. [17] also involved “representatives of families with affected children” in their discussions about outcomes in clinical trials of children with bipolar affective disorder. Pavletic et al. [13], at the end of their report, acknowledge “patients and patient and research advocacy groups.” The level of involvement of these people was not described in detail in any of these articles.

**Industry and drug regulatory authority representatives.** Three groups (Carlson et al. [17], Ramsey et al. [19], and the NDDI [20–24]) specifically mention that representatives from industry or the Food and Drug Administration (FDA) were present. The NDDI is described as a collaboration between the FDA and “neonatal experts and colleagues, representing industry and academia” [10]. Carlson describes invited participants in the group selecting outcomes for clinical trials of children with bipolar disorder as including “pharmaceutical industry sponsors with an interest in mood stabilizer products, staff of the FDA and their counterparts from regulatory agencies in Canada and the European Union” [17]. The Cystic Fibrosis Foundation sponsored a consensus conference that also included “representatives from both the Cystic Fibrosis Foundation and the U.S. Food and Drug Administration” [12].

**Techniques Used to Validate Outcomes**

Three groups made some attempt to validate the outcomes they had selected.

Gianinni et al. [14] assessed the multicollinearity and redundancy of a core set of outcomes for use in clinical trials of children with rheumatoid arthritis by measuring them in a group of children in a clinical practice setting, and using a database from a previous observational cohort study. The acceptability of the core set of outcomes to a wider group of clinicians was assessed by sending a questionnaire to an international selection of rheumatologists seeking their reactions to the outcomes.

The IMACS group retrospectively assessed the validity,
reliability, and responsiveness of the outcomes they had selected by reviewing available literature on the topic [15].

The PRINTO group prospectively validated the core set of outcomes they had produced for clinical trials of children with SLE [30]. This was done by measuring the outcomes in patients in a clinical out patients setting who were being started on new modalities of treatment for their condition. In this way the authors aimed to “mirror” a clinical trial setting. The feasibility, discriminative ability, validity, and internal consistency of the core set of outcomes were assessed in this way.

All three of these groups also developed “definitions of improvement,” based on the degree of change within each outcome, which could be used as a dichotomous index in clinical trials to determine whether patients had benefited from the treatment they had received. This was done in all cases by developing a set of “paper patient profiles,” and asking a group of experts whether or not they thought the patient had improved. A set of potential definitions of improvement was then narrowed down to a final definition by way of consensus formation techniques.

Which Outcomes Were Selected by the Groups?

In Table S5 we summarise the outcomes that were selected by each group, categorised into the following outcome domains: disease activity; disease complications; adverse effects of therapy; functional status; social outcomes, family outcomes and Quality of Life; resource utilisation.

Discussion

To our knowledge this is the first systematic review of studies that addressed selection of outcomes for use in clinical trials in children.

We identified 13 groups formed to address the issue of selecting outcomes for use in paediatric clinical trials. Certain groups—notably, those who have selected outcomes for clinical trials of children with rheumatological conditions—have specifically highlighted that it is inappropriate simply to use the outcomes utilised in adult clinical trials in a paediatric population.

We identified three methods used for reaching consensus, namely NGT, Delphi technique, and semistructured discussion. Many groups used a multidisciplinary approach to the problem of outcome selection, including researchers with experience of clinical trial design, statisticians, and clinicians. Some groups also involved representatives from industry or drug regulatory authorities, but the nature of their involvement is not evident from reading the reports.

No group among the studies we reviewed directly involved children in the process of selecting outcomes. As the aim of clinical trials should be to determine whether patients experience important benefits from an intervention, it was notable that we did not identify any studies that had directly asked children what they considered to be the most relevant outcome domains or outcomes. In the United Kingdom steps are being taken to involve consumers in medical research. A major initiative is the James Lind Alliance (http://www.lindalliance.org), a collaboration with the aim of ascertaining from patients what they think are the most pressing research priorities for various conditions. Determining appropriate outcomes for paediatric studies is thus another area in which consumer involvement in clinical trial design should be encouraged. The difficulties of undertaking this task, however, should not be underestimated [31].

Robustness and Limitations of the Review

Our review was conducted in a rigorous, systematic manner. Two reviewers adhered to strict eligibility criteria to determine which studies should be included. Although the sample of excluded papers checked by the second reviewer represented a small proportion of all the ineligible studies, we concluded that agreement between the reviewers was adequate. We determined that a smaller proportion of excluded studies would be sufficient for quality assurance as compared to a review in which we were meta-analysing the results of clinical trials; possible missed studies were considered an acceptable tradeoff.

There were recurring features of the methodology and reporting quality of the consensus statements that may have compromised the scientific validity of the studies we identified. Most studies that described formation of a consensus statement did not explain in sufficient detail two key aspects of the process—namely, the method used to select group participants, and the process by which consensus was reached. Insufficient information was given to determine the level of involvement of certain groups involved in the research, particularly industry representatives, drug regulatory authority representatives, and parents of affected children.

Although the 8,889 abstracts identified were screened twice, it may be possible that some relevant studies were missed. The types of studies that may not have been identified at this stage include clinical trials that did not describe in the abstract how the authors selected their outcomes, but subsequently in the full text may have mentioned the process used. It is also possible that some studies may have been missed by not searching the “grey” literature such as unpublished conference proceedings.

We excluded studies that did not state specifically that they selected outcomes for use specifically in clinical trials in children. Our reason for this exclusion was that such studies should involve patients themselves, and the unique challenges of doing this in children warrant the separation of adult and paediatric studies. Another group of studies excluded were those concerned with the development of assessment tools for outcomes such as quality of life. Although this work is crucial for designing valid assessment tools, and will to some degree ascertain from children how illness affects their life, these studies focussed on how to measure an outcome rather than what outcome to measure.

Another set of studies outside the scope of this review were those relating to the selection of outcomes that are measured in newborns as a surrogate measure of maternity care given to women. For example, one way of evaluating the efficacy of antenatal care is to measure outcomes in babies such as rates of neonatal infection [32]. Similarly, studies in which outcomes were selected that evaluate the effect of interventions given to children by measuring effects on the family were not systematically sought. We did, however, identify two studies in which such outcomes were selected [19,33].

Core Outcomes

If implemented, the studies we have identified should reduce the impact of inappropriate outcome selection on the
quality of the evidence provided by individual clinical trials. The development of a universally agreed core set of outcomes for a condition could, as well as improving the quality of individual clinical trials, lead to less heterogeneity between trials. One problem associated with nonuniform reporting of outcomes is outcome reporting bias, a phenomenon that results from the selective reporting of some outcomes but not others, depending on the results [34,35].

One cause of outcome reporting bias may be that statistically insignificant results are more likely to be left out of the report, so outcomes at the planning stage of a trial that might otherwise have been deemed clinically relevant are deemed “irrelevant” after data analysis, rendering the published literature a biased and selective representation of the research [36]. Disease-specific, universally agreed core sets of outcomes that should be measured and reported in all clinical trials [36]. Disease-specific, universally agreed core sets of outcomes that should be measured and reported in all clinical trials of a specific condition, regardless of statistical significance, have been advocated as a solution to this common problem [35]. Uniform selection of outcomes would also make interpretation of results and comparison across trials simpler, hence making meta-analyses easier and more powerful [9].

European Drug Regulation

It is increasingly recognized that there is a need for high-quality paediatric clinical trials, and the development of Paediatric Investigation Plans (PIPs) is one of the changes in drug regulation in Europe that should facilitate this goal. The PIP is a detailed outline of the research, submitted to the European Medicines Agency (EMEA), that would be needed to investigate the potential benefits and harms of medications for use in children. If a drug company were to be involved in the writing and implementation of a PIP, they would be eligible for marketing rewards in the form of prolonged patent protection and market exclusivity. When a PIP is submitted, the endpoints selected for the trial must be clearly stated and their appropriateness described (http://www.emea.europa.eu/htms/human/paediatrics/pips.htm). The studies we have identified that suggest to trialists which outcomes to measure should be of use to people designing a PIP, and it is possible that the types of studies we have identified may become more popular as drug companies seek to take advantage of the benefits of conducting high-quality clinical trials.

The new standards for conducting clinical trials of investigational medicinal products set by the EMEA aim to improve the quality of paediatric research. In order to obtain a license for a drug, it must be investigated according to these guidelines http://www.emea.europa.eu/htms/human/humanguidelines/efficacy.htm. In July 2007, of the 13 paediatric conditions identified in this review the EMEA Web site included guidelines for one (juvenile idiopathic arthritis) and a concept paper discussing the need for guidelines for another (cystic fibrosis).

The Selection of Outcomes for Use in Children

It is appropriate that some aspects of study design in clinical trials in children differ from equivalent studies performed in adults, and selection of outcomes is one such issue that trialists should consider. In certain situations outcome selection may be similar between the two groups, and some outcomes could be appropriately transposed from adult studies into trials in paediatric populations either in their original state or with slight modification. The danger, however, of not acknowledging the differences between children and adults with the same disease is that the overall validity of trial results could be compromised. Griffiths et al. [11], when discussing which outcomes to measure in clinical trials of children with Crohn disease, highlight that the importance of linear growth is “unique to pediatric patients.” Another example of an outcome exclusive to children is the assessment of neurodevelopment. Other differences between adults and children that may preclude the use of some outcomes are, both groups include distinct disease pathogenesis, different clinical features and natural history, variations in physiological and psychological outcomes, and contrasting roles within the contexts of families and society in general that may preclude the use of the same outcomes.

Unanswered Questions

The best strategy for selecting outcomes for clinical trials in children is currently not known, and future research in this area is warranted. One important question relates to the involvement of children and parents in the formulation of consensus statements. It seems logical that their involvement would help determine the most appropriate outcomes to measure, but there is no evidence to substantiate this hypothesis, nor is there a framework that could recommend the best strategy for involvement. Another area for research is the investigation of the relative strengths and weaknesses of the consensus formation techniques identified here when applied to the problem of selecting outcomes for paediatric studies.

In summary, we have reviewed studies that address the process of selecting outcomes for clinical trials in children. Although it is commendable that there are existing collaborations in several clinical areas, future work in this area may be improved by involving children and parents in the process. The studies identified by this review will go some way to improving the quality of paediatric research, but further research is justified and urgently needed.

Conclusions

Implications for the practice of designing clinical trials. We identified 13 paediatric conditions for which work has been done to determine which outcomes should be measured in clinical trials. When designing clinical trials in these conditions, this work should make the selection of outcomes easier and more uniform.

Implications for research. Although some work on how to select outcomes in paediatric trials has been published in a few clinical areas, there is a need for similar work to be conducted in other areas. Very little work has been done that involves parents or children in assessing which outcomes should be measured in clinical trials; future research should be undertaken to address this deficiency.

Supporting Information

Table S1. Search Methods for Identification of Studies

Table S2. Characteristics of Excluded Studies

Table S3. Characteristics of Included Studies

Found at doi:10.1371/journal.pmed.0050096.s001 (114 KB DOC).

Found at doi:10.1371/journal.pmed.0050096.s002 (243 KB DOC).

Found at doi:10.1371/journal.pmed.0050096.s003 (145 KB DOC).
Table S4. Critical Appraisal of the Methodological Quality of Included Studies

Table S5. Outcomes Selected for Use in Clinical Trials in Children

Text S1. QUOROM Checklist

Acknowledgments

We are grateful to Dr. Tony Marson and Dr. Bridget Young, who provided helpful comments during this review. We would also like to thank Miss Natalie Yates for her help in searching the medical databases. We are also grateful to the members of the MCRN Clinical Study Groups who helped us identify relevant studies and to Miss Jennifer Blakeburn for her assistance in contacting these members.

Author contributions. IS, RLS, and PRW designed the study protocol and search strategy. IS and LJ identified the relevant studies from the search results. IS and LJ extracted data, which were checked by PRW. IS, RLS, and PRW were involved in data analysis. IS prepared the initial manuscript. IS, RLS, and PRW were all substantially involved in the revision of this manuscript. All authors contributed to the final manuscript before submission.

References

Editors’ Summary

**Background.** When adult patients are given a drug for a disease by their doctors, they can be sure that its benefits and harms will have been carefully studied in clinical trials. Clinical researchers will have asked how well the drug does when compared to other drugs by giving groups of patients the various treatments and determining several “outcomes.” These are measurements carefully chosen in advance by clinical experts that ensure that trials provide as much information as possible about how effectively a drug deals with a specific disease and whether it has any other effects on patients’ health and daily life. The situation is very different, however, for pediatric (child) patients. About three-quarters of the drugs given to children are “off-label”—they have not been specifically tested in children. The assumption used to be that children are just small people who can safely take drugs tested in adults provided the dose is scaled down. However, it is now known that children’s bodies handle many drugs differently from adult bodies and that a safe dose for an adult can sometimes kill a child even after scaling down for body size. Consequently, regulatory bodies in the US, Europe, and elsewhere now require clinical trials to be done in children and drugs for pediatric use to be specifically licensed.

**Why Was This Study Done?** Because children are not small adults, the methodology used to design trials involving children needs to be adapted from that used to design trials in adult patients. In particular, the process of selecting the outcomes to include in pediatric trials needs to take into account the differences between adults and children. For example, because children’s brains are still developing, it may be important to include outcome measures that will detect any effect that drugs have on intellectual development. In this study, therefore, the researchers undertook a systematic review of the medical literature to discover how much is known about the best way to select outcomes in clinical trials in children.

**What Did the Researchers Do and Find?** The researchers used a predefined search strategy to identify all the studies published since 1950 that examined the selection of outcomes in clinical trials in children. They also asked experts in pediatric clinical research for details of relevant studies. Only 25 studies, which covered several pediatric specialties and were published by 13 collaborative groups, met the strict eligibility criteria laid down by the researchers for their systematic review. Several approaches previously used to choose outcomes in clinical trials in adults were used in these studies to select outcomes. Two groups used the “Delphi” technique, in which opinions are sought from individuals, collated, and fed back to the individuals to generate discussion and a final, consensus agreement. One group used the “nominal group technique,” which involves the use of structured face-to-face discussions to develop a solution to a problem followed by a vote. Another group used both methods. The remaining groups (except one that used a questionnaire) used semistructured discussion meetings or workshops to decide on outcomes. Although most of the groups included clinical experts, people doing research on the specific clinical condition under investigation, and industry representatives, only three groups asked parents about which outcomes should be included in the trials, and none asked children directly.

**What Do These Findings Mean?** These findings indicate that very few studies have addressed the selection of appropriate outcomes for clinical research in children. Indeed, in many pediatric specialties no research has been done on this important topic. Importantly, some of the studies included in this systematic review clearly show that it is inappropriate to use the outcomes used in adult clinical trials in pediatric populations. Overall, although the studies identified in this review provide some useful information on the selection of outcomes in clinical trials in children, further research is urgently needed to ensure that this process is made easier and more uniform. In particular, much more research must be done to determine the best way to involve children and their parents in the selection of outcomes.

**Additional Information.** Please access these Web sites via the online version of this summary at http://dx.doi.org/10.1371/journal.pmed.0050096.

- A related PLoS Medicine Perspective article is available
- The European Medicines Agency provides information about the regulation of medicines for children in Europe
- The US Food and Drug Administration Office of Pediatric Therapeutics provides similar information for the US
- The UK Medicines and Healthcare products Regulatory Agency also provides information on why medicines need to be tested in children
- The UK Medicines for Children Research Network aims to facilitate the conduct of clinical trials of medicines for children
- The James Lind Alliance has been established in the UK to increase patient involvement in medical research issues such as outcome selection in clinical trials
Outcomes in Clinical Trials of Inhaled Corticosteroids for Children with Asthma Are Narrowly Focussed on Short Term Disease Activity

Ian P. Sinha1*, Paula R. Williamson2, Rosalind L. Smyth1

1 Institute of Child Health, University of Liverpool, Liverpool, United Kingdom, 2 Centre for Medical Statistics and Health Evaluation, University of Liverpool, Liverpool, United Kingdom

Abstract

Background: Little work has been done to determine which outcomes should be measured in randomised controlled trials (RCTs) in children with asthma. Drug regulatory authorities require that short term disease activity is measured, but other outcome domains are not mandatory for licensing and marketing purposes. We aimed to identify whether any domains were underrepresented in RCTs of regular therapies for children with asthma over a 20 year period, and to examine what consistency there was between RCTs in the outcomes used to assess the domains.

Methodology/Principal Findings: By searching the Cochrane Central Register of Controlled Trials in January 2008, we identified all parallel-group RCTs, published between January 1988 and December 2007, which assessed inhaled corticosteroids (ICS) as regular therapy for children with asthma. We evaluated how frequently RCTs measured the following pre-defined domains: disease activity; disease damage; functional status; quality of life; health resource utilisation; and adverse effects of therapy. Our initial search identified 1668 abstracts, of which 412 were retrieved in full. 159 RCTs, of which 115 involved only children and 44 involved children and adults, were included in the review. Disease activity was measured in 157 RCTs, adverse effects of ICS in 135, functional status in 25, quality of life in 21, and health resource utilisation in 17. No RCT measured long term disease damage, although two used FEV1 as a measure of ‘lung growth’. RCTs were inconsistent in the outcomes used to measure the domains.

Conclusions: Short term disease activity is the most frequently measured outcome domain in RCTs in children with asthma. Effects of regular therapies on functional status, quality of life, and long term consequences of asthma are infrequently assessed. A core set of outcomes, developed using consensus techniques, would standardise the measurement of appropriate outcomes in these RCTs. Involving patients would identify outcomes which are most relevant from their perspective.

Introduction

Asthma in children is a major global health problem [1], because it is an important cause of morbidity, mortality and economic cost [2], it is the commonest chronic condition in industrialised countries [3], its prevalence is increasing [4], and in many children it is a progressive condition that continues into adulthood [5].

The first line of regular therapy for the control of asthma in children is inhaled corticosteroids (ICS), and recommended additional treatments are long acting beta-2 agonists and leukotriene receptor antagonists [6]. Randomised controlled trials (RCTs) are the most scientifically rigorous method for evaluating the efficacy and safety of these medications [7], but it can be difficult to select the most appropriate outcomes to measure in these studies, because asthma impacts on many aspects of the lives of children. For example, the effects of treatment could include improvement of daily symptoms, quality of life, or physiological tests of lung function such as Forced Expiratory Volume in 1 second (FEV1) or Peak Expiratory Flow Rate (PEFR) [8].

We previously published a systematic review, of studies that determined which outcomes should be measured in clinical trials in children [9]. In this review, we proposed that outcomes measured in RCTs that include children should be considered under six domains: short term measures of disease activity, physical consequence of disease, functional status, family outcomes and quality of life, side effects of therapy and health resource utilisation. We found one study that addressed the outcomes used in clinical trials of regular therapies for childhood asthma, in which the authors ascertained, by questionnaire, the opinions of 14 specialists and researchers about outcomes relating to disease activity, functional status, and quality of life [10]. As this study...
consulted only a limited group of experts, and did not use recognised consensus techniques, it does not provide a robust basis for recommendations about which domains and outcomes are most appropriate for RCTs of children with asthma.

There have been initiatives to standardise the outcomes which are measured in clinical trials of other conditions. The most notable is the OMERACT collaboration, an international network of clinicians and patients, initially formed in response to the observation that clinical trials of patients with rheumatoid arthritis conducted in the USA measured different outcomes to those conducted in Europe. OMERACT uses structured consensus techniques to determine which outcomes should be measured in clinical trials in a variety of rheumatological conditions [11–13]. Initiatives such as these increase the likelihood that all important outcome domains are measured, reduce the measurement of inappropriate outcomes [14], and aid comparison and synthesis of findings between different clinical trials [15,16]. It has also been suggested that arbitrary or inconsistent outcome selection may lead to clinical trials with unnecessarily large sample sizes [17] and reporting biases [17,18].

The aim of this systematic review was to assess which outcomes had been measured in clinical trials of ICS in children with chronic asthma between 1988 and 2007, in order to determine whether all relevant domains were represented, and whether there was consistent selection of outcomes within these domains. Secondary objectives were to determine whether the selection of outcome domains has changed between 1988 and 2007, whether domains measured in RCTs exclusively involving children differ from those in studies involving both children and adults, and whether domains measured in publically funded trials differed from those in trials funded by the pharmaceutical industry.

Methods

Included studies

In order to ensure that we assessed a group of similar studies, we limited this review to include only RCTs with parallel group design that assessed ICS as a therapy to prevent symptoms or long-term effects of asthma in children. We excluded crossover trials because they are generally of acute interventions, the length of treatment in these studies is typically shorter, and the outcomes they measure may differ from those measured in parallel trials [19]. In order to only include RCTs assessing long term preventative therapy for asthma, we excluded studies with a treatment phase of less than one month. The review was restricted to studies published between January 1988 and December 2007.

Identification of studies

Using the abbreviated search strategy ‘children AND inhaled corticosteroids AND asthma’, the Cochrane Central Register of Controlled Trials was searched in January 2008. This database comprises RCTs from MEDLINE, and also from conference proceedings and journals not indexed in MEDLINE. The references of identified studies were also screened for other potentially eligible studies. The full search strategy is included in supplementary File S1. One reviewer (IS) assessed trial eligibility, under the supervision of the senior authors (PRW and RLS).

Data extraction and quality assessment

The same reviewer extracted the following data, and any problems were resolved by discussion with the other two authors:

1. All outcomes measured in the trial
2. If stated, the designated primary outcome, and whether this was described in sufficient detail, including the methods used to measure it, by whom, and when it was measured and analysed.

3. Masking of interventions was examined because the extent to which interventions were masked may affect the choice of outcomes, and whether they were measured objectively or subjectively. The adequacy of masking was categorised as follows.

   Adequately masked. Authors either clearly describe or imply, in the methods, how the allocated treatment was masked to the patient and family, medical caregiver, and relevant trial personnel involved in measuring outcomes.

   Inadequately masked. Authors specifically state that the identity of the allocated treatment arm was not masked to at least one of the following: patient and family, medical caregiver, or relevant trial personnel involved in measuring outcomes.

   Unclear. Unclear from study methods whether masking was adequate or not.

4. Other study features: year of publication; interventions compared; ages of children included; length of study treatment; source of funding; single- or multi-centre study

Data analysis and presentation

For each study that included exclusively children, the data were tabulated and each outcome was grouped into one of the following six outcome domains, some of which were further divided into subdomains [9]: disease activity, physical consequence of disease, functional status, social outcomes and quality of life, side effects of therapy and health resource utilisation. Where it was unclear which domain was appropriate, this was resolved by discussion between the authors.

To assess how the selection of outcomes has changed over time, we divided the period 1988 to 2007 into sixteen separate epochs, each lasting five years. In each epoch we calculated the proportion of studies measuring each outcome domain, and we presented the results as a moving window.

Results

Flow of included studies

The search yielded 1668 potentially eligible reports. Of these, 1256 were excluded by reading the abstract. The remaining 412 were retrieved in full, and 203/412 were subsequently excluded. In total, 209 eligible reports, of 159 RCTs, were included in the review. The review flowchart is shown in Figure 1. Included studies are listed in supplementary File S2.

Description of included studies

Of the 159 studies included in this review, 115 exclusively included children, and 44 included children and adults.

The characteristics of included studies are summarised in Table 1. Within the group of studies that included only children, all paediatric age groups were represented, but only 25/115 (21%) included children younger than four years of age. In the studies of adults and children, 42/44 (95%) included children aged between 12 and 18 years of age, but not younger than 12, and 2/44 (5%) included children between 5 and 18 years of age, but not younger than 5.

83/159 (52%) included a comparison between ICS groups (either different doses, modes of delivery, or different types of ICS eg fluticasone vs beclomethasone), 63/159 (40%) included a comparison with placebo, and 54/159 (34%) included a comparison with another drug.

Masking of interventions was deemed adequate in 121/159 studies (76%), inadequate in 33/159 (21%), and unclear in 5/159 (3%). Of the 33 studies that were classified as inadequately masked, 18 compared ICS with another drug, 8 compared ICS administered by different devices, 3 compared one ICS to another,
3 compared ICS administered using different dosing schedules, and 1 compared ICS with no treatment. Subjective outcomes that could have been affected by the lack of blinding were measured in 29/33 of these studies (29/33 measured symptoms, 8/33 measured quality of life and 6/33 measured functional status).

### Outcome domains which were measured in the studies

Disease activity was measured in 157/159 (99%) studies, adverse effects of therapy in 135/159 (85%), functional status in 25/159 (16%), quality of life in 21/159 (13%), and health resource utilisation in 17/159 (11%). No studies measured the effects of ICS on long-term physical consequences of asthma, although two studies measured post-bronchodilator FEV1, as a percentage of the predicted value, to assess 'lung growth'. In one of these studies children, aged between 5 and 12 years, were randomised to receive inhaled budesonide, nedocromil sodium or placebo for a period of four to six years, and FEV1 was measured as the primary outcome [20]. In the other study, patients aged between 5 and 66 years were randomised to treatment with inhaled budesonide or placebo for three years, and FEV1 was measured as a secondary outcome [21]. Similar outcome domains were represented in the 115 studies that included only children and the 44 that included children and adults.

There was a wide variety of outcomes within individual domains. This was greatest for the disease activity domain, which was divided into five subdomains (clinical measures, physiological tests of lung function, global measures, bronchial responsiveness to a challenge agent or exercise and markers of inflammation), each of which included outcomes measured in different ways. As can be seen in Table 2, the selection of subdomains and outcomes was inconsistent across the studies.

### Primary outcomes

The primary outcomes measured in the RCTs that included only children are listed in Table 2. It was possible to determine the primary outcome in 84/115 (73%) studies. In 64 of these, the primary outcome was clearly stated by the authors, and in the remaining 20 it was inferred from the outcome used to calculate the sample size. Five studies each selected two co-primary outcomes. Of the 94 studies that specified a primary outcome, 74 (80%) selected primary outcomes that measured disease activity. A total of 17 different primary outcomes were selected, of which physiological measures of airway obstruction, including PEFR (26 studies) and FEV1 (16 studies), were the most frequent. None of the primary outcomes addressed the functional status or quality of life domains.

It was possible to determine the primary outcome in 39/44 (89%) studies that included children and adults. In 34 of these, the primary outcome was clearly stated by the authors, and in the remaining 5 it was inferred from the outcome used to calculate the required sample size. In 38/39 (97%), was some measure of disease activity. The most widely used primary outcome was FEV1 (28 studies).

### Outcomes measured in studies funded by the pharmaceutical industry

The frequency with which most domains were measured in the 127 studies sponsored by the pharmaceutical industry was similar to the 32 publicly funded studies. The main difference we observed was that adverse effects of therapy were measured in a higher proportion of studies sponsored by the pharmaceutical industry (118/127, 93%) compared to studies funded from other sources (17/32, 53%).

### How the selection of outcome domains has changed over time

The trend over the period January 1988 to December 2007 in the selection of outcome domains in RCTs including only children is shown in Figure 2. Disease activity and adverse effects of therapy have remained consistently frequently measured outcome domains. Since the 1992–1996 epoch the proportion of studies measuring functional status, for example by assessing school absence due to asthma, has decreased from 40% to 10%, and those measuring quality of life have increased from 10% to 25%.

### Discussion

We found that RCTs in children with asthma almost always assess the effects of therapies on short term disease activity, but none consider the effects on long term progression of disease. Quality of life and functional status are measured infrequently. While there were similarities between studies, particularly in the selection of primary outcomes that measure disease activity, other outcomes showed wide variability.

The pharmaceutical industry funded 80% of the RCTS we identified, and so it is not surprising that the frequency with which outcomes in the disease activity domain have been measured as primary or secondary outcomes reflects, to some extent, the requirements of the FDA [22,23] and EMEA [24–27]. These authorities recommend, for the purpose of drug licensing and marketing authorisation, that risks and benefits of preventative therapies for children with asthma are assessed in clinical trials that measure, as primary outcomes, physiological tests of pulmonary function and clinical measures such as symptom scores. Other measures of short term disease activity, such as use of rescue medication, rate of exacerbations, and bronchial hyper-respon-
siveness are suggested as important outcomes. Quality of life and exercise tolerance are mentioned as additional outcomes that may provide useful information, but no clear recommendations have been made regarding the use of these outcomes [25].

It is disappointing that, despite the use of ICS in childhood asthma for more than twenty years, their effects on functional status, quality of life, and long term consequences of asthma remain largely unknown. Markers of short-term disease activity, despite their prominence in drug regulatory guidelines and popularity amongst trialists, have been shown to correlate poorly with quality of life [28–30], and are therefore not appropriate surrogate markers for aspects of asthma that could be of more relevance to patients.

Given that the aim of an RCT is to evaluate the safety and efficacy of interventions, and provide some assessment of whether the intervention does more good than harm, it is disappointing to note that quality of life and family outcomes are only measured in 20% of RCTs, and that the impact of disease on functional status is now measured in less than 10%. We identified no clinical trials in which the primary outcome measured these domains. In studies that have investigated which outcomes are clinically relevant to patients with other conditions, such as chronic pain [31], fibromyalgia [32], and rheumatoid arthritis [33], measures of functional status and quality of life were identified as being of great importance, and it is likely that this is the case in children with asthma.

Functional status overlaps with quality of life and disease activity. However, we feel that measures of functional status are important, distinct, markers of how asthma affects children. In clinical trials in adults with chronic illnesses, absence from work is an important outcome, and we feel that an appropriate childhood equivalent would include measures of school attendance and other activities of daily living.

It is particularly important in trials of children to assess the impact of treatments in the long term. Very few studies have

<table>
<thead>
<tr>
<th>Table 1. Characteristics of included studies.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study characteristic</td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>Date of publication</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Length of treatment period</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Age groups of children included</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Number of centres</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Source of funding</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Comparisons</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pone.0006276.t001
Table 2. Frequency with which outcome domains, and outcomes used to measure them, were selected in 115 trials involving only children published between 1988 and 2008.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Subdomain 1</th>
<th>Subdomain 2</th>
<th>Outcome</th>
<th>Number (%) of studies in which measured as primary or secondary outcome n = 115</th>
<th>Number (%) of studies in which measured as primary outcome n = 84*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease activity</td>
<td>Clinical measures n = 109</td>
<td>Symptoms</td>
<td>Symptom severity</td>
<td>114 (99)</td>
<td>74 (88)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Symptom frequency</td>
<td>90 (77)</td>
<td>10 (12)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Use of rescue therapy</td>
<td>90 (77)</td>
<td>2 (2)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Exacerbations</td>
<td>Exacerbation frequency</td>
<td>35 (30)</td>
<td>4 (5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Time to exacerbation</td>
<td>10 (9)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Tests of lung function n = 103</td>
<td>Spirometry</td>
<td>FEV1</td>
<td>80 (70)</td>
<td>16 (19)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>FVC</td>
<td>31 (26)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mid expiratory flow</td>
<td>23 (20)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>FEV1:FVC</td>
<td>6 (5)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>FEV1 reversibility</td>
<td>9 (8)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PEFR</td>
<td>PEFR</td>
<td>85 (73)</td>
<td>26 (31)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Diurnal variability</td>
<td>13 (11)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Day-to-day variability</td>
<td>5 (4)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lung volume</td>
<td>Plethysmographic</td>
<td>4 (3)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Airway flow</td>
<td>Resistance/conductance</td>
<td>5 (4)</td>
<td>2 (2)</td>
<td></td>
</tr>
<tr>
<td>Global measure of control n = 29</td>
<td>Physician-rated</td>
<td></td>
<td>8 (7)</td>
<td>1 (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Parent/patient – rated</td>
<td></td>
<td>14 (12)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>‘Treatment failure’</td>
<td></td>
<td>13 (11)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>‘Treatment success’</td>
<td></td>
<td>3 (3)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Bronchial responsiveness to a challenge agent n = 29</td>
<td>Induced BHR</td>
<td>Methacholine-induced</td>
<td>26 (22)</td>
<td>6 (7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Exercise-induced</td>
<td></td>
<td>7 (6)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Exhaled nitric oxide</td>
<td></td>
<td>5 (4)</td>
<td>1 (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Leukotriene5/interleukin4</td>
<td></td>
<td>4 (3)</td>
<td>1 (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Eosinophils6/IgE7</td>
<td></td>
<td>18 (15)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Physical consequence of disease</td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>HRU</td>
<td>Unscheduled HRU</td>
<td></td>
<td>15 (13)</td>
<td>1 (1)</td>
<td></td>
</tr>
<tr>
<td>Functional status</td>
<td>Effect of asthma on ADL</td>
<td></td>
<td>10 (9)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>School attendance</td>
<td></td>
<td>15 (13)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>QoL/family outcomes</td>
<td>Child’s QoL</td>
<td></td>
<td>19 (17)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Caregiver QoL</td>
<td></td>
<td>5 (4)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Caregiver functional status</td>
<td></td>
<td>8 (7)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Adverse Effects of therapy</td>
<td>Routinely monitored AE n = 82</td>
<td>Patient/parent- reported</td>
<td>80 (70)</td>
<td>1 (1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Routine laboratory AE</td>
<td></td>
<td>32 (27)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Oropharyngeal infection</td>
<td></td>
<td>28 (24)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ophthalmological events</td>
<td></td>
<td>7 (6)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Urine/serum cortisol</td>
<td></td>
<td>52 (44)</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ACTH stimulation</td>
<td></td>
<td>17 (15)</td>
<td>1 (1)</td>
<td></td>
</tr>
</tbody>
</table>
attempted to study the impact of ICS on modifying or affecting the physical consequences of asthma. Two studies, the Childhood Asthma Management Program (CAMP) [20], and the Inhaled Steroid Treatment as Regular Therapy in Early Asthma (START) [21] indicated in their aims that they wished to investigate this effect. Both measured primary outcomes which we have classified as related to disease activity, although the CAMP study stated that their primary outcome (FEV1, %predicted) was a measure of ‘lung growth’. Investigators acknowledge the difficulties in assessing the impact of disease, or therapy, on ‘lung growth’ in children with asthma. It is unlikely that, in a clinical trial, a single measure will provide the best primary outcome and more methodological research is needed to identify whether longitudinal outcomes, such as rate of change in lung function measures, would be a more appropriate way of assessing lung growth. Although 24 of the studies that exclusively included children had a treatment period lasting longer than one year, only three measured outcomes after the end of the treatment period [20,34,35]. The other 21 studies represent missed opportunities to investigate the long term effects of ICS on progression of asthma in children, and this question should be addressed in future clinical trials.

The long term safety of treatments for asthma has recently been identified as being of particular importance to patients and clinicians [36]. Although 83% of studies that only included children assessed the safety of ICS, the quality with which long term systemic side effects of ICS were measured was variable. Of the 24 studies that lasted longer than a year, 21 measured effects on growth. None of the studies measured the final adult height attained, despite the fact that this is of most interest to children and parents. Only 14 studies measured the effect of ICS, when administered for longer than one year, on hypothalamic-pituitary-adrenal function, despite the serious and potentially fatal consequences of this type of adverse reaction [37]. We suggest that serious systemic side effects should be monitored in all clinical trials of ICS in children with asthma, so that both the benefits and risks of these drugs can be appropriately evaluated.

As well as the fact that measurement of long-term efficacy and safety outcomes is not a requirement of drug regulatory authorities, there are other reasons why they may not have been measured in the studies we identified. Diagnostic and technical problems associated with measurement of lung growth, financial cost, and problems with patient attrition also hinder the conduct of long-term studies in children with asthma. Agreement that long-term outcomes are important, amongst clinicians, patients, and researchers, could promote the conduct of such studies. There is also a need for research to identify the most appropriate long-term outcomes, and the ways in which they should be measured.

The studies we identified are comparable in terms of the population they include and the interventions which they compare, and so we feel that our finding of heterogeneity of outcome selection between studies is valid. Even though we have reviewed RCTs assessing one aspect of the treatment of childhood asthma, it is likely that our findings would be similar if we were to conduct a similar review of, for example, clinical trials of leukotriene antagonists or long acting beta2 agonists.

We have reviewed outcomes which have been reported rather than those which were actually measured. Outcome reporting bias in published RCT reports is common [18,38], and in order to have evaluated exactly which outcomes had been measured it may have been more accurate to assess trial protocols. Although outcome reporting bias may lead to an underestimation of the frequency with which some outcomes were actually measured, it is unlikely that it would affect heterogeneity between studies.

Non-uniform outcome selection can make it difficult to design, interpret, and meta-analyse clinical trials [15,16,39,40], and so a few collaborations have begun to address the problems of which outcomes to measure in clinical trials of a variety of paediatric and adult conditions [9,13]. One solution is a universally-agreed core set of outcomes which should be measured, as a minimum, in all clinical trials of a specific condition. Core sets were first designed by the OMERACT group, which utilises structured consensus techniques amongst a diverse group of stakeholders and consumers. Our findings would suggest that a similar initiative in childhood asthma, with separate consideration of pre-school and older children, would make an important contribution to improving clinical research in this very prevalent, chronic disease.
Conclusions

We have shown that outcomes in RCTs in children with asthma, mainly driven by the requirements of drug regulatory authorities, are focussed on short term disease activity, and those which may be more relevant to patients are largely overlooked. Future research must be directed towards determining the most appropriate and important outcomes to measure in these trials.

Supporting Information

File S1  Search strategy
Found at: doi:10.1371/journal.pone.0006276.s001 (0.04 MB PDF)

File S2  List of included studies
Found at: doi:10.1371/journal.pone.0006276.s002 (0.32 MB PDF)
Acknowledgments

We would like to thank Professor Doug Altman for helpful comments during the design of the study protocol.

References


Core Outcome Sets

The Delphi Technique as a Method of Developing

Useful?

What Are Core Outcome Sets and Why Are They

Ian P. Sinha

Review of Existing Studies

Outcomes to Measure in Clinical Trials: Using the Delphi Technique to Determine Which

Guidelines and Guidance

Firstly, researchers can select outcomes that suit their needs, at the expense of outcomes that are of most importance to patients or clinicians [1–3]. Secondly, heterogenous selection and measurement of outcomes in clinical trials can impair the ability to synthesise results across studies in systematic reviews [4]. Thirdly, in the absence of a set of outcomes that should be measured and reported in all clinical trials in the same condition, it can be difficult to ascertain, in the final publication, whether authors report all results or only those that they find favourable [5,6].

As a result, the standardisation of outcomes for clinical trials has been proposed as a solution to the problems of inappropriate and non-uniform outcome selection [4,7] and reporting bias [5,8]. The most notable work relating to outcome standardisation has been conducted by the OMERACT (Outcome Measures in Rheumatology) collaboration, which advocates the use of core outcome sets designed using consensus techniques that are then measured and reported in clinical trials in rheumatology [9]. However, such initiatives are uncommon. In some specialties, such as paediatrics, the number of conditions covered is low and the quality of existing studies variable [10]. In addition, there is limited guidance in the literature regarding the development of a core outcome set. This paper aims to contribute to the methodology of determining which outcomes to measure in clinical trials, or systematic reviews of clinical trials.

The Delphi Technique as a Method of Developing Core Outcome Sets

One method for reaching consensus around which outcomes to measure is the Delphi technique, which comprises sequential questionnaires answered anonymously by a panel of participants with relevant expertise. After each questionnaire, the group response is fed back to participants [11]. In terms of the overall validity of the final consensus, this approach has advantages over less structured methods of reaching consensus such as round-table discussions. Participants in a Delphi study do not interact directly with each other, so situations where the group is dominated by the views of certain individuals can be avoided. When participants consider whether to change their opinion or stick to their original answers, after seeing the group response this decision is not affected by the desire to be seen to agree with senior, overly vocal, or domineering individuals. Improvements in global communication have made it feasible to use the Delphi technique to involve geographically distant participants in larger numbers than are traditionally used in studies employing face-to-face discussion, and so it is also increasingly being used to reach consensus around many topics in medicine, such as education, development of clinical guidelines, and prioritisation of research topics.

There is little guidance for researchers who wish to use the Delphi technique, even though aspects of its methodology can be interpreted in a variety of ways. Most published work has provided guidance based on authors’ experiences, rather than empirical research or theoretical justification for the methodological decisions made. One systematic review describes a variety of consensus techniques used for designing clinical guidelines [12]. The authors highlighted important methodological decisions that may affect the overall quality of the final consensus, such as the types of participants involved, the questions they are asked, the information they receive to inform their answers, the manner of the interaction between them, and the way in which consensus is agreed. These have also been variously highlighted as important aspects of methodology in other commentaries about the Delphi technique [13–15].

To our knowledge, there is no guidance related to methodological considerations or reporting for studies using the Delphi technique to determine which outcomes or domains to measure in clinical research studies. The objective of the systematic review summarised below (and included in full in Text S1) was to examine studies that used the Delphi technique for this purpose. Our recommendations from this review are then summarised to help inform the conduct and reporting of future initiatives.


Published January 25, 2011

Copyright: © 2011 Sinha et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: IPS was funded by the NIHR Medicines for Children Research Network Clinical Trials Unit and Co-ordinating Centre. The Medicines for Children Research Network is part of the National Institute for Health Research (NIHR), and is funded by the Department of Health. IPS was funded by Department of Health grant RNC/013/011. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Competing Interests: RLS is a member of the PLoS Board of Directors.

* E-mail: iansinha@liv.ac.uk

Provenance: Not commissioned; externally peer reviewed.
Summary Points

- Studies that use the Delphi process for gaining consensus around a core outcome set for clinical trials should be of sufficiently high quality in order for their recommendations to be considered valid.
- We report a systematic review of 15 studies that used the Delphi technique for this purpose, in which we identified variability in methodology and reporting.
- To improve the quality of studies that use the Delphi process for developing core outcome sets, we recommend that patients and clinicians be involved, researchers and facilitators avoid imposing their views on participants, and attrition of participants be minimised.
- Methodological decisions should be clearly described in the main publication in order to enable appraisal of the study.

A Systematic Review of Studies That Have Used the Delphi Technique to Identify Which Outcomes to Measure in Clinical Trials

We searched Medline (no date restrictions) in January 2010 to identify studies that used the Delphi technique to determine which outcomes to measure in clinical trials or systematic reviews of clinical trials. From each eligible study, the following methodological aspects were noted: the participants involved, the types of questions asked, whether the study was completely anonymised, whether non-responders in earlier rounds were included or excluded from subsequent rounds, and the definition of consensus used by the authors. We also evaluated the quality with which the methods and results were reported. These assessments enabled us to identify variations in the methods applied within these studies, and areas of reporting quality that could be improved.

Of 636 abstracts, 20 full text articles were retrieved, of which five were excluded because they aimed to identify outcomes for use in clinical practice, and the authors did not state whether the participants considered their use in clinical research studies. Many of the 636 studies excluded on the basis of the abstract described the use of the Delphi process to develop clinical guidelines and educational curricula. Of 15 studies included in the review, eight developed core outcome sets for rheumatological conditions. Others identified outcomes for pain in children, degenerative ataxia, gastro oesophageal reflux disease, infantile spasms, maternity care, multiple sclerosis, and thyroid eye disease.

Studies varied in terms of group composition and the manner in which the Delphi process was conducted. Participation in such studies was dominated by researchers, with patients and families seldom involved.

The reporting quality of studies also varied. Important methodological aspects that were generally less well reported were the information provided to participants at the start of the Delphi process, the information fed back to participants after each round, and the level of anonymity. A summary of the reporting quality of the studies is shown in Table 1. Each of the items included in the table had been highlighted, by one or more of the commentaries mentioned earlier [13–15], as an important methodological consideration when using the Delphi technique. We tailored the

Table 1. Reporting quality of the 15 included studies.

<table>
<thead>
<tr>
<th>Broad Aspect of Reporting</th>
<th>Specific Items for Which the Reporting Quality Was Assessed</th>
<th>Studies in Which Clearly Reported</th>
<th>Studies in Which Not Clearly Reported</th>
<th>N/A</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Size and composition of the panel</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of participants</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Types of participants (e.g., clinicians, patients)</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Proportion of each type of participant</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>How participants were identified/sampled</td>
<td>14</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>Methodology of the Delphi process</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Administration of questionnaires (e.g., postal)</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>How items were generated for first questionnaire</td>
<td>14</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>What was asked in each round</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Information provided to participants before the first round</td>
<td>6</td>
<td>9</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>How the overall group response was fed back to participants</td>
<td>8</td>
<td>7</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Level of anonymity (total or quasi-anonymity)</td>
<td>4</td>
<td>11</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>A priori definition of “consensus” about whether an outcome should be measured</td>
<td>7</td>
<td>1</td>
<td>7*</td>
<td></td>
</tr>
<tr>
<td>Were non-responders invited to subsequent rounds</td>
<td>10</td>
<td>0</td>
<td>5*</td>
<td></td>
</tr>
<tr>
<td><strong>Results</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of respondents to each round</td>
<td>14</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Number who completed every round</td>
<td>11</td>
<td>4</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Results for each outcome in each round</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Group response for each outcome (final round)</td>
<td>8</td>
<td>7</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Distribution of response for each outcome in the final round</td>
<td>7</td>
<td>8</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>List of all outcomes that participants agreed should be measured</td>
<td>8</td>
<td>0</td>
<td>7*</td>
<td></td>
</tr>
</tbody>
</table>

*Reaching a final consensus was not the aim of the Delphi process, so a definition of consensus was not given.

*All participants responded to each round, so no discussion was made regarding non-responders.

do:10.1371/journal.pmed.1000393.t001
statements so they were relevant for the Delphi process as a method of developing consensus around a core outcome set.

Although an assessment of response rate to each round could be made in 14/15 studies, it was only possible to accurately assess attrition rates in 11/15 studies, which reported the proportion of first round respondents who also completed the final round. Of these, only six studies reported the proportion of participants who completed every round in the Delphi process, from start to finish. Only seven reports presented a measure and distribution of the group opinion for each outcome listed in the final round. No study reported the results, in each round, for every outcome that was considered by the group.

**Guidance about Using the Delphi Technique to Determine Core Outcome Sets**

**Involve Clinicians and Patients**

Informed clinical decisions can only be based on the results of trials that have measured outcomes of importance to both clinicians and patients. Initiatives to identify which outcomes to measure in clinical trials, however, focus on the opinions of researchers. This means that outcomes included in existing core sets may be selected to serve the needs of researchers in academia or industry, rather than considering how important they are to patients.

Patients have a variety of perspectives about living with a condition, which may differ from those of clinicians and researchers. In one study, involvement of patients in the design of a systematic review highlighted certain outcomes as being of particular importance, but these had not been measured in any of the included trials [16]. Research conducted within the OMERACT group also suggests that clinicians and researchers may not realise that certain outcomes are very important for patients [17]. The perspective of patients is now routinely incorporated into the work conducted by OMERACT [18]. Another important initiative, which actively promotes the involvement of patients and families in identifying priorities in clinical research, is the James Lind Alliance (http://www.lindalliance.org/). In a recent systematic review, this group found a few examples of conditions that should be reported in all studies that use the Delphi technique [25]. A recommended checklist of study characteristics and results considered by all stakeholders [1,2,21].

**Try to Minimise Attrition**

People with minority opinions may be more likely to drop out of studies that use the Delphi process, so attrition as rounds progress can lead to overestimation of the degree of consensus in the final results. Strategies to prevent attrition bias are to only invite people who respond to a pre-Delphi invitation to participate in the first round [22] or to list, in the publication, only those participants who either completed the entire Delphi process, or agreed the final consensus statement [23]. An example of a paragraph that could be used to explain to participants the importance of completing the whole Delphi process is shown in Box 1.

**Report Certain Aspects of the Methodology and Results**

In order to enable appraisal of the quality of studies that use the Delphi process to identify outcomes that should be measured in clinical research, which may in turn affect whether the recommendations are implemented, authors should describe certain important methodological features in the study report. Criticisms of the Delphi technique are that “expertise” of the panel is arbitrarily defined, and that the validity of the final consensus is questionable because individual participants are not accountable for their responses, and they may be led towards conformity with the group, rather than consensus of true opinions [24]. As described earlier, attrition of participants may mean the degree of consensus reached in the final round is overestimated [25]. A recommended checklist of study characteristics and results that should be reported in all studies that use the Delphi technique to determine which outcomes to measure in clinical research studies is shown in Table 2. Given the variation across previous studies, it would be helpful if authors explained their methodological choices, and discussed the effects these may have on the results.

**Box 1. Example Text to Emphasize to Participants the Importance of Completing the Whole Delphi Process**

Thank you for agreeing to participate in our study. It is very important that you complete the questionnaires in each round. The reliability of the results could be compromised if people drop out of the study before it is completed, because they feel that the rest of the group does not share their opinions. If people drop out because they feel their opinions are in the minority, the final results will overestimate how much the sample of participants agreed on this topic.
Future Areas of Methodological Research

Given variations in methodology between studies, we feel there is a need for research to determine how best to develop core outcome sets. An agenda for this research could be designed through the COMET Initiative (Core Outcome Measures for Effectiveness Trials), which is an international network of individuals and organisations with interest or experience in the development, application, and promotion of core outcome sets (http://www.liv.ac.uk/nwhtmr/comet/comet.htm). One such area of ongoing research and discussion relates to whether core outcome sets designed for clinical practice, such as those developed in the five studies we excluded, should be the same as those designed for research. Another priority is research to identify the most effective ways to incorporate the views of different groups of participants, especially patients, in the design of core outcome sets.

Supporting Information

Text S1 Full report (and PRISMA checklist) of the systematic review of studies that used the Delphi technique to determine which outcomes to measure in clinical trials. Found at: doi:10.1371/journal.pmed.1000393.s001 (0.87 MB DOC)

Acknowledgments

We acknowledge the support of the National Institute for Health Research, through the MCRN.
Author Contributions

ICMJE criteria for authorship read and met: IPS RLS PRW. Agree with the manuscript’s results and conclusions: IPS RLS PRW. Designed the experiments/the study: IPS RLS PRW. Analyzed the data: IPS PRW.

References