Developing the agenda for core outcome set development

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

Elizabeth Gargon

February 2016

Abstract

Author: Elizabeth Gargon
Thesis title: Developing the agenda for core outcome set development

Introduction and aims

A core outcome set (COS) is defined as an agreed standardised set of outcomes that should be measured and reported, as a minimum, in all clinical trials in specific areas of health or health care. Their use allows research to be compared and combined as appropriate, and may ensure that all studies provide usable information. There is currently no accepted gold standard method for COS development and further work was necessary to explore choices about methods, and what the priorities are for guidance and further research in this area. This thesis aimed to investigate what is currently known about COS development, and explore developers' experiences of developing COS.

Methods

A systematic review of studies reporting the development of a COS was undertaken, and the methodological techniques used in these studies was described. A mixed methods approach was undertaken to explore COS development, drawing on qualitative interviews with, and an online web-based survey of, COS developers. This thesis used a Triangulation Design to obtain different but complementary data on the same topic for comprehensiveness.

Results

The systematic review identified 198 published studies that described the development of COS for clinical trials. The systematic review demonstrated variability in the ways that COS had been developed, particularly the methods used and the stakeholders included as participants in the process. Patient participants had infrequently been included in the development of COS (18%). Key aspects of the process were frequently not reported.

Eighty-one (48%) developers completed the survey. The majority of survey respondents (73%) felt that there is a need for methodological guidance or research to inform future activity to develop COS. Areas for future guidance or research included: stakeholder involvement, patient involvement in particular; choice of methodology, and consensus formation.

32 interviews were conducted with COS developers (18 with published, and 14 with ongoing, COS projects). Developers found the process of COS development to be a challenging process, in part due to the nature of COS development being an emerging field of research, but also in part to not always considering important methodological details from the outset, for example their choice of methods and stakeholders. There was a variety of influences on developers' choice of methods, which included the previous literature on COS development, expert advice, developers' own experience with methods and the resources available to developers. The absence of guidance in COS development, and the prominence of uncertainties, dominated developers' accounts.

Conclusions

The work in this thesis has brought COS together in one place for the first time, summarises key characteristics of COS and their development, and provides the first comprehensive account of COS development. It will inform the development of much needed guidance in this area and help to improve COS development methodology. Guidance needs to determine commonalities across different disease areas, and promote awareness of important issues; encourage COS developers to think about their own contexts and circumstances, and enable COS developers to make decisions about methods that best suit their needs and resources. Guidance seems to be needed for all aspects of COS development, but it was particularly felt that guidance around the systematic review process, conduct of Delphi, and conduct of consensus meetings, are high priority.

Acknowledgements

I would like to thank my supervisors, Professors Paula Williamson and Bridget Young, for their continued support and guidance. I would also like to thank my many colleagues in Biostatistics for their unwavering encouragement.

Several people helped enormously with the work contained in this thesis, and deserve my thanks: Shona Kirtley, for her comments when developing the search strategy for the systematic review, outlined in Chapter 2; Binu Gurung and Nancy Medley, for their extreme hard work and commitment as my co-reviewers during the systematic review; and the COMET Management Group for their overall help and support, particularly Mike Clarke whose knowledge and insight is an inspiration.

I am grateful to the MRC North West Hub for Trials Methodology Research (Ref G0800792 and MR/K025635/1) for funding this PhD and giving me this opportunity.

I thank my dear father for everything that I am and always believing in me. I thank my partner, Simon, for his constant love, friendship and support, without which I would be lost. Finally, I dedicate this thesis to my mum, my guardian angel, for always looking out for me.

Statement of contribution

The systematic review reported in this thesis formed part of the Core Outcome Measures in Effectiveness Trials (COMET) Initiative. The COMET Management Group consists of Doug Altman (University of Oxford), Jane Blazeby (University of Bristol), Mike Clarke (Queen's University Belfast), Paula Williamson (University of Liverpool) and myself as the Project Coordinator. The Management Group provided comments on the search strategy (presented in Chapter 3 and Appendix 1) and all read and approved the final manuscript for publication. Mike Clarke provided additional comments on Chapter 2 of this thesis, and read and approved the final manuscript for publication.

Shona Kirtley (University of Oxford) provided comments on the technical aspects of the MEDLINE search strategy (presented in Chapter 3 and Appendix 1), as well as its modification for use in other databases.

Binu Gurung and Nancy Medley were second reviewers for the systematic review presented in Chapter 3 of this thesis. They were involved in the process of selecting studies for inclusion in the review, checking for correct exclusion in the review, and data extraction.

Ethics statement

Ethics approval was not required for the work in Chapters 1 to 5 of this thesis.

The work described in Chapters 6-7 fully complied with the ethical practice guidelines laid out by the British Psychology Association. Approval was sought and granted through the University of Liverpool's ethics procedures (Research Ethics Subcommittee for Non-Invasive Procedures reference number: RETH000624).

Table of contents

List of tables

Number *Page*

List of figures

Number *Page*

Abbreviations used in this thesis

CDE	Common data elements
CDISC	Clinical Data Interchange Standards Consortium
CFAST	Coalition for Accelerating Standards and Therapies
CINAHL	Cumulative Index of Nursing and Allied Health Literature
CMR	Cochrane Methodology Register
COMET	Core Outcome Measures in Effectiveness Trials
CONSORT	Consolidated Standards of Reporting Trials
COS	Core outcome set(s)
COSMIN	The COnsensus-based Standards for the selection of health Measurement Instruments
CROWN	Core Outcomes in Women's Health
COA	Clinical Outcome Assessments
CRG	Cochrane Review Group
EQUATOR	Enhancing the QUAlity and Transparency Of health Research
FDA	Food and Drug Administration
GRADE	Grading of Recommendations Assessment, Development and Evaluation
HOME	Harmonising Outcome Measures for Eczema
HTA	Health Technology Assessment
ICF	International Classification of Functioning, Disability and Health
ICHOM	International Consortium for Health Outcomes Measurement
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
IMMPACT	Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials
MRC	Medical Research Council
NDDI	Neonatal Drug Development Initiative
NICE	National Institute for Health and Care Excellence
NIH	National Institutes of Health
NIHR	National Institute for Health Research
NLM	National Library of Medicine
NNR	Number Needed to Read
NWHTMR	North West Hub for Trials Methodology
OMF	Outcomes Measures Framework
OMERACT	Outcome Measures in Rheumatology
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses

PROMIS	Patient Reported Outcomes Measurement Information System
RCT	Randomised Controlled Trial
SDO	Standards Development Organization
SEALD	Study Endpoints and Labeling Development
SPIRIT	Standard Protocol Items: Recommendations for Interventional Trials
WHO	World Health Organisation

Initials of researchers involved in the work in this thesis:

BG	Binu Gurung
BY	Bridget Young
EG	Elizabeth Gargon
NM	Nancy Medley
PW	Paula Williamson

Publications and presentations

Chapter 1

Sections of the Introduction chapter will be published in Journal of Comparative Effectiveness Trials:

Tunis, S, M Clarke, S Gorst, **E Gargon**, et al. (2016). "Improving the Relevance and Consistency of Outcomes in Comparative Effectiveness Research." Journal of Comparative Effectiveness Research (*Article in press*).

Chapter 2

The work contained in chapter 2 has been published in BMC Medical Research Methodology:

Gargon, E, PR Williamson, et al. (2015). "Collating the knowledge base for core outcome set development: developing and appraising the search strategy for a systematic review." BMC Med Res Methodology 15(1): 26. *Highly accessed.*

This work was presented in a Cochrane Evidence Podcast for International Clinical Trials Day 2015: http://www.cochrane.org/podcasts/ICT-day-2015/core-outcomes

Chapter 3

The work contained in chapter 3 has been published in PLoS One:

Gargon, E, B Gurung, et al. (2014). "Choosing important health outcomes for comparative effectiveness research: a systematic review." PLoS ONE **9**(6): e99111.

The work in this chapter was presented at the following conferences:

Gargon, E, B Gurung, et al. (2013). Collating the knowledge base for the COMET (core outcome measures in effectiveness trials) initiative - a systematic review. 2nd Clinical Trials Methodology Conference: Methodology Matters. Edinburgh, 18-19 November 2013. Trials 2013, **14** (Suppl 1):O67 doi:10.1186/1745-6215-14-S1-O67 (Oral)

Gargon, E, B Gurung, et al. (2014). Choosing important health outcomes for comparative effectiveness research: a systematic review. ISPOR 17th Annual European Congress. Amsterdam, 10-12 November 2014 (Poster)

Gargon, E, D Altman, et al. (2015). The need for standardised outcomes in cancer clinical trials: a report of cancer core outcome sets. National Cancer Intelligence Network Cancer Outcomes Conference 2015. Belfast, 8-10 June 2015. Abstracts of oral presentations. European Journal of Cancer Care, 24: 1–23. doi: 10.1111/ecc.12329 (Oral)

Chapters 5 to 7

The work in these chapters was presented at the following conferences:

Gargon, E, B Young, et al. (2014). A survey of core outcome set developers. The 4th Meeting of the COMET Initiative. Rome, 19-20 November 2014. Trials 2015, 16(Suppl 1):O1 doi:10.1186/1745-6215-16-S1-O1 (Oral)

Gargon, E, B Young, et al. (2015). A mixed methods study of researchers' experiences of developing core outcome sets. ISPOR 18th Annual European Congress. Milan, 9-11 November 2015 (Oral and poster)

Gargon, E, B Young, et al. (2015). A mixed methods study of researchers' experiences of developing core outcome sets. 3rd International Clinical Trials Methodology Conference 2015. Glasgow, 16-17 November 2015 (Poster)

A copy of the publications arising from the work in Chapters 2 and 3 is included in Appendix 16.

The following publications are also referred to in the thesis:

Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, **Gargon E,** Tugwell P. (2012) Developing core outcome sets for clinical trials: issues to consider. Trials 13: 132.

Kirkham JJ, **Gargon E**, Clarke M, Williamson PR (2013) Can a core outcome set improve the quality of systematic reviews?--a survey of the Co-ordinating Editors of Cochrane Review Groups. Trials 14: 21.

Gargon E, PR Williamson, et al. (2014). "The COMET Initiative database: progress and activities from 2011 to 2013." Trials **15**(1): 279.

Smith V, Clarke M, Williamson P, **Gargon E** (2014) Survey of new 2007 and 2011 Cochrane reviews found 37% of prespecified outcomes not reported. Journal of Clinical Epidemiology.

Gargon E, Williamson PR, Altman DG, Blazeby JM, Clarke M. The COMET initiative database: progress and activities update (2014). Trials 2015;16(1):515.

Kirkham JJ, Gorst S, Altman DG, Blazeby J, Clarke M, Devane D, **Gargon E**, Williamson PR. (2015) COS-STAR: a reporting guideline for studies developing core outcome sets (protocol). Trials 16: 373.

Gorst, SL, **E Gargon**, M Clarke, JM Blazeby, et al., Choosing Important Health Outcomes for Comparative Effectiveness Research: An Updated Review and User Survey. PLoS ONE, 2016. 11(1): p. e0146444.

Chapter 1: Introduction

1.1 Clinical trials

Clinical trials are research studies undertaken for the purpose of assessing the safety and efficacy of interventions, treatments or care procedures. The focus of the work in this thesis is clinical trials with human beings. Randomised controlled trials (RCTs) are seen as the gold standard in evaluating the effects of treatments [1], largely due to the randomisation of treatment allocation in an RCT. This prevents selection bias by distributing the characteristics of patients that may influence assessment of treatment between groups, hence allowing an unbiased assessment of treatment effect [2]. Clinical trial data has many uses, including to inform clinical guidelines and shared decision making practices; labeling to provide information that is most useful to prescribes in treating patients [3], and in the development of health policies such as those by The National Institute for Health and Clinical Excellence (NICE) [4].

Researchers, clinicians and policymakers often distinguish between the *efficacy* and the *effectiveness* of an intervention. Whereas efficacy trials (also described as explanatory trials) determine whether an intervention can have a beneficial effect in an ideal situation under optimum conditions [5], effectiveness trials (also described as pragmatic trials) measure the degree of beneficial effect under "real world" clinical settings. In contrast to an efficacy trial, an effectiveness trial will usually be conducted following as close to clinical practice as possible [6]. Design of effectiveness trials are therefore based on conditions of, and with consideration to, routine clinical practice and clinical decision making. Efficacy trials tend to precede effectiveness trials, and although it is preferential to distinguish between efficacy and effectiveness trials, in reality they exist on a continuum [1], often making it difficult to separate the two as distinct phases of research. The focus of this thesis hereon in will be effectiveness trials.

1.2 Outcomes in trials

There are three basic components of RCTs [7]:

1. at least one test treatment and a comparator treatment;

2. randomisation of persons to treatment,
3. outcome measure(s).

It is the third component that is the focus of this thesis. Broadly, in the context of clinical trials, an outcome is defined to be a measurement or observation used to capture and assess the effect of treatment, such as assessment of side effects (risk) or benefits. When designing a clinical trial, the PICO format is often used to formulate a research question. A 'well-built' question should include four parts; that is identifying the patient problem or population (P), the intervention (I), the comparator (C) and the outcomes of interest (O) [8]. In a randomised trial, differences between the groups in outcomes can be inferred to be as a result of the differing interventions. Therefore the selection, measurement and reporting of important, relevant and appropriate outcomes are critical.

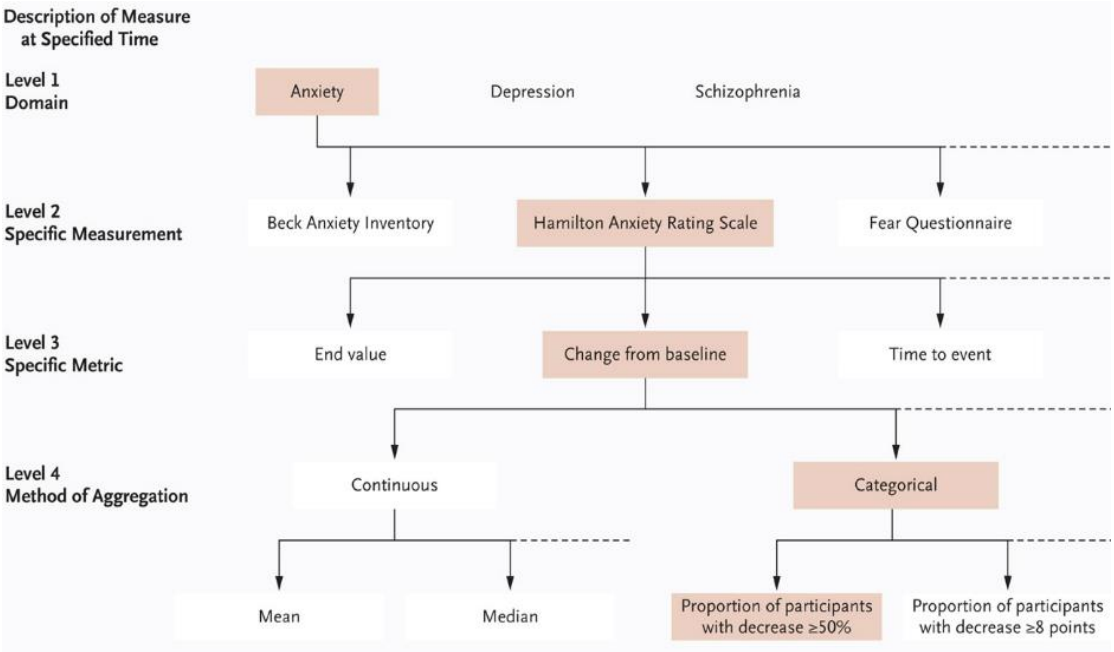
Clinical trials will usually include multiple outcomes of interest, and the main outcomes are usually those essential for decision making. Some outcomes will be of more interest than others. The primary outcome should be the outcome deemed most capable of providing clinically relevant evidence directly related to the objective of the trial [9]. The primary outcome is the outcome considered to be of greatest importance to relevant stakeholders (such as patients, clinicians, policy makers, funders, researchers). It often represents the greatest therapeutic benefit [10], is an integral component of the research question under investigation and is usually the one used in the sample size calculation [11]. Sometimes researchers propose more than one primary outcome if they are thought to be of equal therapeutic importance and relevance to the research question. This can also be useful if it is unclear which single primary outcome will best answer the question. However, consideration should be given to the effect on the type 1 error (the incorrect rejection of a true null hypothesis) because of the potential for multiplicity problems, for example the extent of intercorrelation among the proposed primary outcomes [9].

Secondary outcomes evaluate other beneficial or harmful effects of secondary importance or are useful for explaining additional effects of the intervention [12]. Secondary outcomes may also be exploratory in nature. Harmful effects should always be viewed as important regardless of their primary or secondary outcome label [11].

A variety of different types of outcomes can be measured in trials, and researchers must decide which of these to measure. As well as the importance of an outcome to relevant stakeholders, researchers must consider an array of information including how responsive it is to the interventions being compared and the appropriateness to the trial, for example the financial cost associated with that outcome. The decision is made more complex by the numerous types of outcomes that exist, and researchers must decide which of these types of outcomes is most appropriate for both the question under investigation and the specific context of the clinical trial. For example, a clinical outcome describes a medical event(s) that occurs as a result of disease or treatment [13], and relates to a patient's symptoms, overall mental state or how the patient functions. In contrast, a surrogate end point is used as a substitute for a clinical outcome [14] and has been defined as a biomarker intended to substitute for a clinical outcome [15]; an example would be prostatic specific antigen (PSA) in prostate cancer. A biomarker is a medical sign, typically used in earlier phase trials, used to predict biological processes. Examples of biomarkers include everything from pulse and blood pressure through basic chemistries to more complex laboratory tests of blood and other tissues [16].

In addition to deciding what to measure, Zarin et al (Figure 1) describe that a fully specified outcome measure includes information about the following [17]: domain (e.g. anxiety), that is what to measure; specific measurement (e.g. Hamilton Anxiety Rating Scale), that is how to measure that outcome/domain; the specific metric used to characterise each participant's results (e.g. change from baseline at specified time), and method of aggregation (e.g. a categorical measure such as proportion of participants with a decrease greater than 50%).

Figure 1: An example of the four levels of specification in reporting outcome measures



Reproduced with permission from Zarin, D. A., T. Tse, et al. (2011). "The ClinicalTrials.gov results database-update and key issues." N Engl J Med 364(9): 852-860., Copyright Massachusetts Medical Society.

Furthermore, outcomes can also be measured in different ways. Some clinical outcomes are composed of a combination of items, and are referred to as composite outcomes. A composite outcome combines two or more components into a single measure [18]. Outcomes can be objective, that is not subject to a large degree of individual interpretation, and these are likely to be reliably measured across patients in a study, by different health care providers, and over time. Laboratory tests may be considered objective measures in most cases. Outcomes may also be considered to be subjective. Most clinical outcomes involve varying degrees of subjectivity, for example a diagnosis or assessment by a health care provider, carer or the patient themselves. A clinician reported outcome is an assessment that is determined by an observer with some recognised professional training that is relevant to the measurement being made. In contrast, an observer reported outcome is an assessment that is determined by an observer who does not have a background of professional training that is relevant to the measurement being made, i.e., a non-clinician observer such as a teacher or caregiver. This type of assessment is often used when the patient is unable to self-report (e.g., infants, young children). Finally, a patient reported outcome is a measurement based on a report that comes

directly from the patient (i.e., the study participant) about the status of particular aspects of, or events related to, a patient's health condition [13].

1.3 Outcome domains

Outcome domains are constructs which can be used to classify broad aspects of the effects of interventions e.g. functional status. Outcomes from multiple domains may be important to measure in trials, and several outcomes within a domain may be relevant or important. Outcome domain models or frameworks exist to attempt to provide essential structure to the conceptualisation of domains [19], and have been used to classify outcomes that have been measured in clinical trials in particular conditions. Despite their intended use to provide a framework, there has not always been consistency between the different models. In a review of Health Related Quality of Life (HRQOL) models, Bakas et al. found that there were wide variations in terminology for analogous HRQOL concepts [19]. There have been several frameworks to classify health, disease and outcomes to date,, some of which are described below.

1.3.1 Outcome-related frameworks

World Health Organisation (WHO)

The WHO definition of health, although strictly a definition of health, can be considered a framework as it includes three broad health domains [20]: physical, mental and social well-being. This definition has not been amended since 1948 but is a useful starting place to study health. In a scoping review of conceptual frameworks, Idzerda et al. point out that although the three domains are clearly outlined, no further information about what should be included within each domain is provided [21].

Patient-Reported Outcomes Measurement Information System (PROMIS)

The PROMIS domain framework builds on the WHO definition of health to provide subordinate domains beneath the broad headings stated above [22]: physical (symptoms and functions), mental (affect, behaviour, and cognition) and social well-being (relationships and function). It was developed for adult and paediatric measures as a way of organising outcome measurement tools.

World Health Organization International Classification of Functioning Disability and Health (WHO ICF)

The International classification of Functioning, Disability and Health (ICF) offers a framework to describe functioning, disability and health in a range of conditions. The ICF focuses on the assessment of an individual's functioning in day-to-day life. It provides a framework for body functions, activity levels and participation levels in basic areas and roles of social life; providing domains of biological, psychological, social and environmental aspects of functioning [23]. In many clinical areas, ICF core sets have been developed. These core sets identify the most relevant ICF domains for a particular health condition. Further discussion about ICF core sets can be found in Chapter 3.

5Ds

The 5Ds is presented as a systematic structure for representation of patient outcomes and includes five 'dimensions': death, discomfort, disability, drug or therapeutic toxicity, and dollar cost [24]. This representation of patient outcome was developed specifically for rheumatic diseases, and the authors claim that each dimension represents a patient outcome directly related to patient welfare; for example they describe a patient with arthritis may want to be alive, free of pain, functioning normally, experiencing minimal side effects and be financially solvent. This framework assumes that outcomes are multidimensional, and it is critical that the 'concept of outcome' is orientated to patient values.

Wilson and Cleary

Wilson and Cleary [25] propose a taxonomy or classification for different measures of health outcome. They suggest that one problem with other models is the lack of specification about how outcomes interrelate. They divide outcomes into five levels: biological and physiological factors, symptoms, functioning, general health perceptions, and overall quality of life. In addition to classifying these outcome measures, they propose specific causal relationships between them that link traditional clinical outcomes to measures of health related quality of life. For example, 'Characteristics of the environment' are related to 'Social and psychological supports' which in turn relates to 'Overall quality of life.' Ferrans and colleagues [26] revised the Wilson and Cleary model to further clarify and develop individual and environmental factors.

Outcome Measures in Rheumatology (OMERACT) Filter 2.0

The OMERACT Filter 2.0 [27] is a conceptual framework that the authors claim encompasses 'the complete content of what is measurable in a trial.' That is, a conceptual framework of measurement of health conditions in the setting of interventions. It comprises three core areas: death, life impact and pathophysiologic manifestations; it also comprises one strongly recommended, resource use. These core areas are then further categorised into core domains. They liken the areas to 'large containers' for the concepts of interests (domains and subdomains). They recommend that the ICF domains are also considered under life impact (ICF domains: activity and participation) and pathophysiologic manifestations (ICF domains: body function and structure).

Outcome Measures Framework (OMF)

The Outcome Measures Framework (OMF) project was funded by the Agency for Healthcare Research and Quality (a branch of the U.S. Department of Health and Human Services) to create a conceptual framework for development of standard outcome measures used in patient registries [28]. The OMF has three top level broad domains: characteristics, treatments and outcomes. There are six subcategories within the outcomes domain: survival; disease response; events of interest; patient/caregiver reported outcomes; clinician reported outcomes, and health system utilisation. The model was designed so that it can be used to define outcome measures in a standard way across medical conditions. Gliklich et al conclude that 'as the availability of healthcare data grows, opportunities to measure outcomes and to use these data to support clinical research and drive process improvement will increase.'

Survey of Cochrane reviews

Rather than attempting to define outcome domains as others have done, Smith et al performed a review of outcomes from Cochrane reviews to see whether there were similar outcomes across different disease categories, in an attempt to manage and organise data [29]. Fifteen categories of outcomes emerged as being prominent across Cochrane Review Groups and encompassed person-level outcomes, resource-based outcomes, and research/study-related outcomes. The 15 categories are: adverse events or effects (AE), mortality/survival, infection, pain, other physiological or clinical, psychosocial, quality of life, activities of daily living (ADL), medication, economic, hospital, operative, compliance (with treatment), withdrawal (from treatment or study), and satisfaction (patient, clinician,

or other health care provider). The authors recognise that these 15 categories might collapse further and could be 'mapped' to the 'core' areas identified by Boers et al. in the OMERACT framework (described above).

1.4 Problems with outcomes

Clinical trials seek to evaluate whether an intervention is effective and safe by comparing the effects of interventions on outcomes, and by measuring differences in patient outcomes between groups. Clinical decisions about the care of individual patients are made on the basis of these outcomes, so clearly the selection of outcomes to be measured and reported in trials is critical. The chosen outcomes need to be relevant to health service users and others involved in making decisions and choices about health care. However, a lack of adequate attention to the choice of outcomes in clinical trials has led to avoidable waste in both the production and reporting of research, and the outcomes included in research have not always been those that patients regard as most important or relevant [30].

It has been widely shown that inconsistencies in outcomes cause problems for people trying to use healthcare research. One such example is a cross-sectional study of oncology research that found that more than 25,000 outcomes had appeared only once or twice in oncology trials [31]. Furthermore, key outcomes may go unmeasured or unreported, and a review of missing data in Cochrane Reviews found that 102/143 (71%) reviews were unable to obtain the findings for key outcomes in the included trials, and 26 (18%) were missing data for more than half the patients on the review's pre-specified primary outcome [32]. There are also often differences in how outcomes are defined and measured making it difficult, sometimes impossible, to synthesise the results of different research studies and apply them in a meaningful way. This was epitomised in a survey of trials involving people with schizophrenia, where it was found that 2194 different scales had been used in 10,000 controlled trials, meaning that, on average a new instrument had been introduced for every fifth trial [33].

Alongside this inconsistency in the measurement of outcomes, outcome reporting bias adds further to the problems faced by users of research. Outcome reporting bias has been defined as the selection of a subset of the original recorded outcomes, on the basis of the results, for inclusion in the published reports of trials and other research [34]. Publication

of complete trial results is important to clinicians, consumers, and policy makers who wish to make well-informed decisions about health care. However, this does not always happen and outcomes that are statistically significant are more likely to be fully reported [35]. Furthermore, in a sensitivity analysis that sought to account for outcome reporting bias in systematic reviews with a statistically significant result, it was found that 19% would not have remained significant and 26% would have overestimated the treatment effect by more than 20% [36]. Selective reporting of outcomes means that fully informed decisions cannot be made about the care of patients, resource allocation, research priorities, and study design. This can lead to the use of ineffective or even harmful interventions, and to the waste of health care resources that are already limited [37].

1.5 Standardising outcomes

1.5.1 Core outcome sets (COS)

These issues of inconsistency and outcome reporting bias could be reduced with the development and application of agreed standardised sets of outcomes, known as core outcome sets (COS), that should be measured and reported in all trials for a specific clinical area [38]. These sets represent the minimum that should be measured and reported in all clinical trials of a specific condition and could also be suitable for use in other types of research and clinical audit [39]. The expectation is that the core outcomes will always be collected and reported and that researchers might also include other outcomes of particular relevance or interest to their specific study. The first step in the process is typically identifying 'what' to measure. Once it has been agreed what should be measured, the 'how' can be determined; that it is how the outcomes included in the core set should be defined and measured, as well as the timing of such measurements.

The use of COS will make it easier for the results of trials to be compared, contrasted and combined as appropriate, thereby reducing waste in research [40]. This approach would reduce heterogeneity between trials because all trials would measure and report the agreed important outcomes, lead to research that is more likely to have measured relevant outcomes due to the involvement of relevant stakeholders in the process of determining what is core, and be of potential value to use in clinical audit. Importantly, it would enhance the value of evidence synthesis by reducing the risk of outcome reporting bias and ensuring that all trials contribute usable information.

1.5.2 COS initiatives

One of the earliest examples of an attempt to standardise outcomes is an initiative by the World Health Organisation (WHO) in the 1970s, relating to cancer trials [41]. More than 30 representatives from groups doing trials in cancer came together, the result of which was a WHO handbook of guidelines recommending the minimum requirements for data collection in cancer trials. The most notable work to date relating to outcome standardisation since has been conducted by the Outcome Measures in Rheumatology (OMERACT) collaboration, which advocates the use of COS, designed using consensus techniques, in clinical trials in rheumatology. This, and other relevant initiatives, is described below.

OMERACT (www.omeract.org) is an independent initiative of international health professionals interested in outcome measures in rheumatology. The first OMERACT conference on rheumatoid arthritis was held in Maastricht, in the Netherlands in 1992 [42]. The motivation for this was discussions between two of the executive members, comparing the outcomes for patients with rheumatoid arthritis in European clinical trials with that of North American clinical trials, and noting that they used different outcomes. This made it extremely difficult to compare and combine in meta-analyses. Over the last 20 years, OMERACT has served a critical role in the development and validation of clinical and radiographic outcome measures in rheumatoid arthritis, osteoarthritis, psoriatic arthritis, fibromyalgia, and other rheumatic diseases. OMERACT strives to improve outcome measurement in Rheumatology through a 'data driven', iterative consensus process involving relevant stakeholder groups [43].

An important aspect of OMERACT now is the integration of patients at each stage of the OMERACT process, but this was not always the case. Initially, OMERACT did not include patients in the process of developing COS. The patient perspective workshop at OMERACT 6 in 2002 addressed the question of looking at outcomes from the patient perspective. Fatigue emerged as a major outcome in rheumatoid arthritis, and it was agreed that this should be considered for inclusion in the core set [44-46]. This patient input along with clinical trialist insight, epidemiologist assessment, and industry perspective, has led OMERACT to be a prominent decision making group in developing outcome measures for all types of clinical trials and observational research in rheumatology. OMERACT have now

developed COS for many rheumatologic conditions, and have described a conceptual framework for developing core sets in rheumatology (described in section 1.3.1) [27].

Since OMERACT there have been other examples of similar COS initiatives to develop recommendations about the outcomes that should be measured in clinical trials. One example is the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT, www.immpact.org), whose aim is to develop consensus reviews and recommendations for improving the design, execution, and interpretation of clinical trials of treatments for pain. The first IMMPACT meeting was held in November 2002, and there have been a total of seventeen consensus meetings on clinical trials of treatments for acute and chronic pain in adults and children. Another exemplar is the Harmonising Outcome Measures for Eczema (HOME, www.nottingham.ac.uk/homeforeczema) Initiative. This is an international group working to develop core outcomes to include in all eczema trials.

1.5.3 The COMET Initiative

The COMET (Core Outcome Measures in Effectiveness Trials) Initiative (www.comet-initiative.org) brings together people interested in the development and application of COS. COMET aims to collate and stimulate relevant resources, both applied and methodological, to facilitate exchange of ideas and information, and to foster methodological research in this area. Specific objectives include:

1. to raise awareness of current problems with outcomes in clinical trials;
2. to encourage COS development and uptake;
3. to promote patient and public involvement in COS development;
4. to provide resources to facilitate these aims;
5. to avoid unnecessary duplication of effort,
6. to encourage evidence-based COS development.

The COMET Initiative was launched at a meeting in Liverpool in January 2010, funded by the Medical Research Council (MRC) North West Hub for Trials Methodology (NWHTMR). More than 110 people attended, with representatives from trialists, systematic reviewers, health service users, clinical teams, journal editors, trial funders, policy makers, trials registries and regulators. The feedback was uniformly supportive, indicating a strong consensus that the time was right for such an initiative. The meeting was followed by a

second meeting in Bristol in July 2011, which reinforced the need for COS across a wide range of areas of health and the role of COMET in helping to coordinate information about these. COMET has gone on to have subsequent successful international meetings in Manchester (2013), Rome (2014) and Calgary (2015) to affirm this.

1.5.4 Other relevant initiatives

While the initiatives described in Section 1.5.2 are specific to the development of COS for trials in particular areas of health, there are a few other recent initiatives relevant to the improvement of outcome measurement. One such initiative is the Core Outcomes in Women's health (CROWN) initiative. CROWN is an international group, led by journal editors, to harmonise outcome reporting in women's health research [47]. This consortium aims to promote COS in the specialty, encourage researchers to develop COS and facilitate reporting of the development of COS.

The International Consortium for Health Outcomes Measurement (ICHOM) organises global teams of physician leaders, outcomes researchers and patient advocates to define core sets of outcomes per medical condition for use in clinical practice rather than clinical trials. Health care is very complex and medical knowledge is changing fast. Reliable outcomes data enable physicians and patients to make better decisions about what treatments are best for them and who should provide them. ICHOM provides a structured process to achieve consensus for a global standard [48]. This is a new initiative, and they aim to publish 50 standard sets by 2017. A list of completed sets, in progress and conditions under consideration can be viewed at: http://www.ichom.org/medical-conditions/. One such example is the recently published localised prostate cancer 'standard set' [49]. The publication provides limited methodological detail that would be valuable to know, such as how patients and other stakeholders were identified and selected, the weighting of those stakeholder groups and the processes of how an outcome is finally included in the standard set [50]. In addition, it is also important to know how the measuring tools were selected and substantiated as multiple definitions exist. Ultimately, the work being done by ICHOM to develop standard sets for clinical care should be complementary to work being done to develop COS for trials and research.

The US National Institutes of Health (NIH) encourages the use of common data elements (CDEs) in NIH supported research projects or registries. The NIH provide a resource portal

(http://www.nlm.nih.gov/cde/) that includes databases and repositories of data elements and case report forms that may assist investigators in identifying and selecting data elements for use in their projects. PROMIS (http://www.nihpromis.org/) is another NIH initiative and is part of the NIH goal to develop systems to support NIH-funded research supported by all of its institutes and centres. PROMIS provides a system of measures of patient-reported health status for physical, mental, and social health which can be used across chronic conditions (see description in section 1.3.1). Once it has been decided what outcomes should be measured, PROMIS is a source of information regarding how those outcomes could be measured.

Once a COS has been agreed, it is then important to determine how the outcomes included in the set should be defined and measured. Several measurement instruments may exist to measure a given outcome, usually with varying psychometric properties (e.g. reliability and validity). Important sources of information for selecting a measurement instrument for a COS are systematic reviews of measurement instruments. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) initiative collates systematic reviews of measurement properties of available measurement instruments that intend to measure (aspects of) health status or (health-related) quality of life. An overview of these reviews and guidelines for performing such reviews can be found on the COSMIN website (http://www.cosmin.nl/systematic-reviews-of-measurement-properties-_5_0.html.) They have also developed a checklist about which measurement properties are important and standards for how to evaluate their measurement properties [51]. The COSMIN checklist will facilitate the selection of the most appropriate PRO measure among competing instruments. A collaboration between COSMIN and COMET has recently resulted in the development of a guideline on how to select outcome measurement instruments for outcomes included in a COS [52].

The Clinical Outcome Assessments (COA) Staff at the US Food and Drug Administration (FDA), previously known as the Study Endpoints and Labeling Development (SEALD) Study Endpoints Team, aim to encourage the development and application of patient-focused endpoint measures in medical product development to describe clinical benefit in labelling. The COA Staff engage with stakeholders to improve clinical outcome measurement standards and policy development, by providing guidance on COA development, validation, and interpretation of clinical benefit endpoints in clinical trials. The FDA defines a COA as a 'measure of patient's symptoms, overall mental state, or the effects of a disease or

condition on how the patient functions.' Put simply, the COA Staff work to ensure that the evidence provided about an outcome instrument can be relied upon in the context of drug development and regulatory decision making.

The Clinical Data Interchange Standards Consortium (CDISC) is a global Standards Development Organization (SDO) 'to develop and support global, platform-independent data standards that enable information system interoperability to improve medical research and related areas of healthcare,' http://www.cdisc.org. CDISC aims to establish worldwide industry standards to support the electronic acquisition, exchange, submission and archiving of clinical research data and metadata to improve data quality and streamline medical and biopharmaceutical product development and research processes. The Coalition for Accelerating Standards and Therapies (CFAST) Initiative is a CDISC partnership, set up to accelerate clinical research and medical product development by creating and maintaining data standards, tools and methods for conducting research in therapeutic areas that are important to public health (http://www.cdisc.org/cfast-0). One of their objectives is to identify common standards for representing clinical data for drug studies in priority therapeutic areas. This includes standardising definitions of outcomes, and the way in which outcomes are described.

1.6 The need for research to improve methodological standards for COS development

1.6.1 Health research guidelines

Science has been described as a 'heterogeneous endeavour' [53], and health research guidelines can help improve the quality of science and research. Over recent years there has been a surge of guidelines to help researchers produce and report better research: "The time has come for all stakeholders to develop and implement policies that increase accessibility of health research, and promote its unbiased translations to the best possible care of patients [37]."

The Enhancing the QUAlity and Transparency Of health Research (EQUATOR) Network was set up to tackle problems of inadequate reporting of health research. The EQUATOR Network website (www.equator-network.org) provides a resource for a collection of guidelines to help researchers publish their research [54]. The Consolidated Standards of

Reporting Trials (CONSORT) statement is one such guideline. It is an evidence based minimum set of recommendations and items for reporting clinical trials [55], and facilitates complete and transparent reporting. It states that trial reports should include a completely defined pre-specified primary and secondary outcome measure(s), including how and when they were assessed. The Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT) initiative is another guideline with the aim of improving the quality of clinical trial protocols [56]. The SPIRIT Statement provides guidance in the form of a checklist of recommended items to include in a clinical trial protocol. It includes a statement encouraging trial investigators to ascertain whether a COS exists relevant to their trial, and if so, to include those outcomes in their trial. Therefore, COS need to be readily found by potential users and be developed in a methodologically rigorous way. Health research guidelines have a greater chance of success if funders support, and journals more actively endorse and implement, these initiatives to improve the usefulness of research [57].

1.6.2 Recent work

Accumulating work in this area has identified the need for general guidance on the development of COS. There is currently little guidance on the development of COS. Sinha and colleagues conducted a systematic review of studies that determined outcomes for paediatric trials [58]. They found that studies were of variable quality, used variable methods and few had involved parents or children in assessing which outcomes should be measured. They concluded that there were recurring features of the methodology and reporting quality of the studies that may have compromised the scientific validity of the studies identified.

They went on to conduct a review of studies that had used the Delphi Technique to determine outcomes for trials [59]. Again, they identified studies of variable methodology and reporting quality. They recommended that “Methodological decisions should be clearly described in the main publication in order to enable appraisal of the study.” Furthermore, they recommended a checklist that should be reported in studies using this method to determine outcomes to measure in trials.

Williamson et al suggest key issues to consider in the development of a COS including its scope, the stakeholder groups to involve, choice of consensus method and the achievement of a consensus [38]. They suggest a checklist of items that groups should

consider when reporting the general development of a COS. Although the issues to consider are described, it is not a formal COS reporting guideline nor is there one available to date.

1.6.3 Rationale for work in this thesis

There is an increasing awareness of the need for greater attention to be given to the outcomes measured in clinical trials, in terms of standardisation and reporting. Furthermore, methodological issues such as the lack of patients in OMERACT originally and their subsequent inclusion leading to the identification of different outcomes, suggests that there is a need for research to determine how best to develop COS. There is currently no accepted gold standard method for COS development and further work is necessary to explore which methods are better or more appropriate than others, and what the priorities are for guidance and further research in this area. As highlighted in this chapter, it is important to identify both what to measure and how to measure outcomes once they have been included in a COS. The focus of this thesis in on the first part of the process, that is identifying what to measure. For COS to be successfully implemented, they need to be easily accessible to researchers and other key groups. They are currently scattered across the health literature. A systematic review of COS is needed to bring these resources together in one place, as well as to elucidate the methods COS developers have used to date.

The focus of the work undertaken for this thesis is to investigate COS development. In particular, the following questions will be explored:

- What is currently known about COS development?
- What influences COS developers' choice of methodology and approach?
- What are the priorities for guidance and further research in this area?

1.7 Structure of this thesis

The remainder of this thesis is structured as follows.

Chapter 2 outlines the methodological approach taken to develop a search strategy for a systematic review of COS. The aim was to compare the contribution of databases to the

identification of included studies, and to find the best combination of methods to retrieve all included studies.

Chapter 3 provides a detailed description of the systematic review methods used and the results. The aim of the systematic review was to identify studies which had the aim of determining which outcomes or domains to measure in all clinical trials in a specific condition, and to identify and describe the methodological techniques used in these studies.

Chapter 4 provides additional information from the systematic review about the COS development methods most commonly used presently in COS development; specifically systematic reviews, the Delphi technique and consensus meetings.

Chapters 5 to 7 present a mixed methods approach to explore COS development.

Chapter 5 focusses on a web based survey conducted to provide quantifiable information about published COS developers' experiences of the COS development process. Contact authors of the COS publications identified in the systematic review were contacted and asked to answer a few short questions about their COS work.

The work described in chapters 6 and 7 involved undertaking in-depth qualitative interviews with COS developers to further methodological understanding of COS development processes. The aim was to generate a detailed description of COS developers' choice of methodological approach, including the factors that have informed the ways in which researchers have developed COS; and to identify priority areas for future methodological research. Interviews were conducted with both published and ongoing COS developers.

Chapter 8 concludes with a summary of the main findings, further developments and recommendations for future work.

1.7.1 Language and style

As a mixed methods thesis, the language and style used throughout this thesis are consistent with the predominant style used for the methodology of the corresponding

chapter. Chapters 2 to 5 are written in the third person; and chapters 6 and 7 use the first person as the dominant style consistent with qualitative inquiry [60].

Chapter 2: Developing an appropriate search strategy

2.1 Background

Clinical trials seek to evaluate whether interventions are effective and safe for patients by comparing their relative effects on outcomes chosen to identify benefits and harms. Decision makers can then use this information to make well-informed healthcare choices. Therefore, it is critical that the outcomes measured and reported in trials are those that are needed by decision makers. However a lack of adequate attention to the choice of outcomes in clinical trials has led to avoidable waste in both the production and reporting of research, and the outcomes included in research have not always been those that patients regard as most important or relevant [30]. The COMET Initiative brings together people interested in the development and application of core outcome sets (COS). COMET aims to collate and stimulate relevant resources, both applied and methodological, to facilitate exchange of ideas and information, and to foster methodological research in this area. The importance of COS is increasingly recognised by research funders. For instance, the National Institute for Health Research's Health Technology Assessment programme in the UK, the Health Research Board in Ireland and the charity Arthritis Research UK, are all highlighting this to researchers seeking funding for new studies. However, the identification of existing COS is not easy.

For core outcome sets to be an effective solution, they need to be easily accessible to researchers and other key groups. They are currently scattered across the health literature, so we have set out to bring these resources together in one place. This is the first known attempt to do this. As part of the COMET Initiative, we are developing a publicly accessible internet-based resource to collate the knowledge base for COS development and the applied work that has already been done according to disease area. This will be a useful resource for trial funders to refer to, for researchers to see what work has been done in their area of interest and for research funders wishing to avoid unnecessary duplication of effort when supporting new activities. It will include planned and ongoing work, as well as published accounts of COS development. Prior to the completion of the systematic review outlined here, 130 relevant studies had been identified through known research networks,

but creating a comprehensive database and keeping the database up to date is key to its value for users and requires a more structured and transparent approach than the ad hoc inclusion of studies as they come to light.

This requires the development and application of an optimal, multi-faceted search strategy to identify work related to the development of COS. It builds on a review of studies that addressed which outcomes to measure in clinical trials in children that was conducted in 2006, which identified work in 17 different paediatric conditions [58]. This, and studies that had been identified in ad hoc ways, was the starting point. However, in order for the database to be comprehensive and up to date, a systematic approach is needed to identify relevant material.

This chapter outlines the methodological approach taken to develop the search strategy for a systematic review to identify studies which sought to determine which outcomes or domains to measure in all clinical trials in a specific condition, and, in turn, to establish a comprehensive database of COS.

2.2 Aims

To develop an appropriate search strategy to identify as many relevant studies as possible within the available resources, and then investigate the performance characteristics of this strategy.

We aimed to compare the contribution of databases towards identifying included studies, and identify the best combination of methods to retrieve all included studies.

2.3 Methods

2.3.1 Developing an appropriate search strategy

We developed a multi-faceted search strategy to search electronic databases using a combination of text words and index terms, adapting the search strategy as appropriate for each database. This process began with an appraisal of the searches from previous reviews of studies that had (a) sought to determine which outcomes to measure in clinical trials in children [58], (b) used the Delphi technique to determine which outcomes to measure in

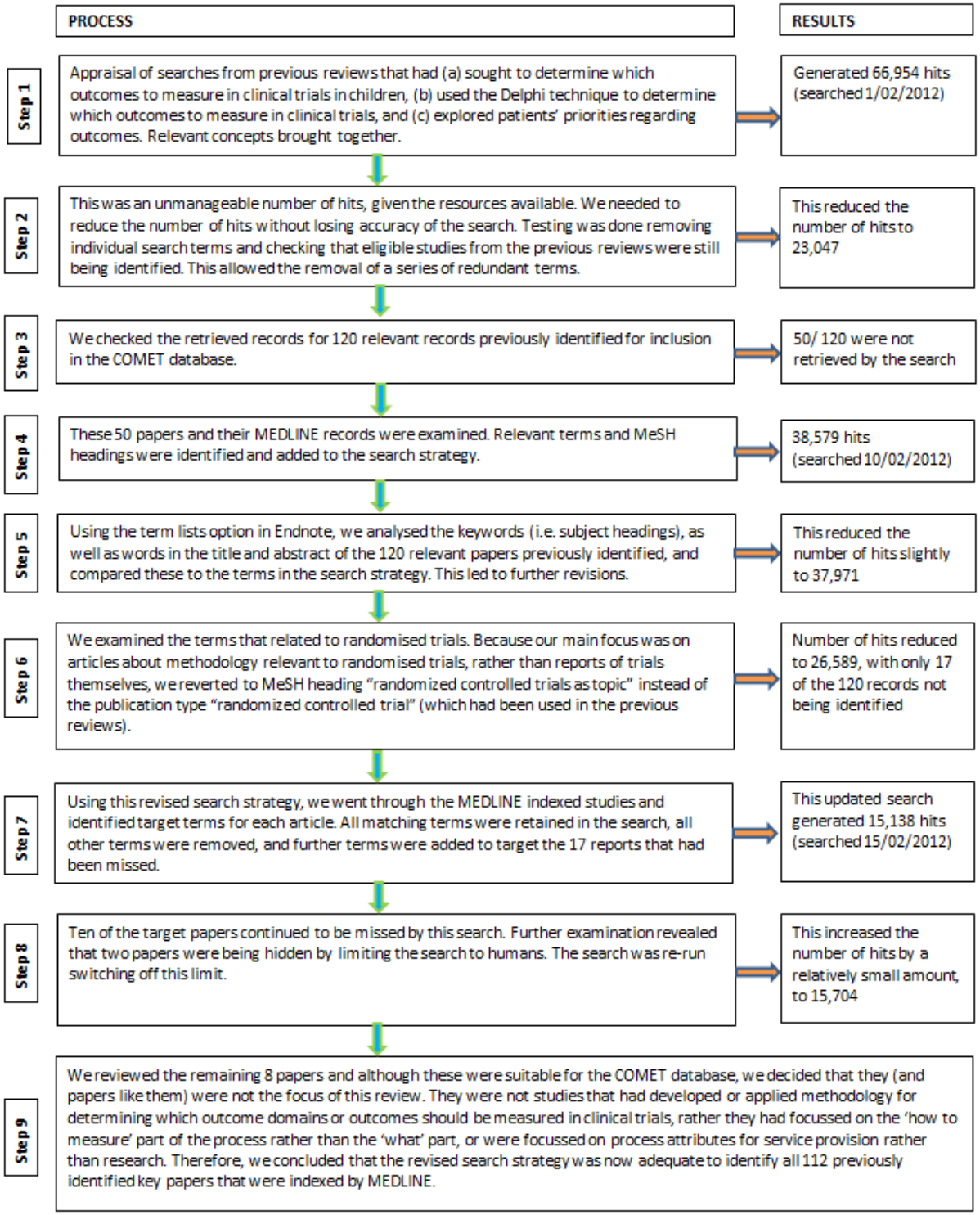
clinical trials [59], and (c) explored patients' priorities regarding outcomes [61]. The relevant concepts from these searches were brought together to develop a search for use in MEDLINE via Ovid. The combined set of search terms generated 66,954 hits (searched 01 February 2012). This was an unmanageable number of hits, given the resources available for the project. We needed to reduce the number of hits, without losing precision of the search. Testing was done, removing individual search terms and checking that eligible studies from the previous reviews were still being identified. This allowed the removal of a series of redundant terms, and reduced the number of hits to 23,047.

The next stage in the development of the search strategy was to compare the retrieved records with the 130 citations that had been identified over time for inclusion in the COMET database of COS. We also wished to retrieve similar reports in the final search strategy and the results of the search for development studies were checked for these 130 citations. Ten of these citations were not indexed for MEDLINE but, of the 120 that were, 50 had not been retrieved by the search. These papers and their MEDLINE records were examined, and terms and MeSH headings that were relevant and would identify those citations were identified. These terms were added to the search strategy, and a re-run in MEDLINE generated 38,579 hits (searched 10 February 2012). Using the *term lists* option in Endnote, we analysed the keywords (i.e. subject headings), as well as words in the title and abstract of the relevant papers previously identified, and compared these to the terms in the search strategy. This led to further revisions, which reduced the number of hits slightly to 37,971.

The next stage in the revision of the search strategy, which had originated from the earlier reviews, was to examine the terms used to identify reports that related to randomised trials. Because our main focus was on articles about methodology relevant to randomised trials, rather than reports of trials themselves, we reverted to MeSH heading "randomized controlled trials as topic" instead of the publication type "randomized controlled trial" (which had been used in the previous reviews). The total number of hits fell to 26,589, with only 17 of the 120 citations that we knew were indexed in MEDLINE not being identified. Using this revised search strategy, we went through the MEDLINE indexed studies and identified target terms for each article. All matching terms were retained in the search, all other terms were removed, and further terms were added to target the 17 reports that had been missed. This updated search generated 15,138 hits (searched 15 February 2012).

However, ten of the target papers continued to be missed by this search. Further examination revealed that two papers were being hidden by limiting the search to humans and switching off this limit increased the number of hits by a relatively small amount, to 15,704. We reviewed the remaining 8 papers [62-69] and although these are suitable for the COMET database, we decided that they (and papers like them) were not of sufficient importance to extend the search further in order to retrieve them. They were not studies that had developed or applied methodology for determining which outcome domains or outcomes should be measured, or are important to measure, in clinical trials. Therefore, we concluded that the revised search strategy was adequate to identify all 112 previously identified key papers that were indexed by MEDLINE. This process is summarised in Figure 2. The search was then modified for use in the other electronic databases that we considered searching for this study. The final search strategies are shown in Appendix 1 and combines three concepts of search terms, covering 'randomised trial / systematic review', 'methodology' and 'outcomes'. All terms within each concept were combined with the Boolean operator OR and the three concepts were then combined using the Boolean operator AND. Key terms were also targeted in the title and abstract fields, and these terms were combined with the Boolean operator OR. Truncation and wildcards were used to improve the sensitivity of the search, account for spelling variations and to identify different derivations of search terms. This search, which was developed for MEDLINE via Ovid, was modified subsequently for use in other electronic databases (see Appendix 1).

Figure 2: Developing an appropriate search strategy



The tables and figures in this chapter are reproduced with permission from Gargon, E., P. R. Williamson, et al. (2015). "Collating the knowledge base for core outcome set development: developing and appraising the search strategy for a systematic review." BMC Med Res Methodol ***15****(1): 26.*

2.3.2 Confirming the need for such a large search

As noted above, the final search was still retrieving a large number of records and our next step was to determine if this was because there were many eligible papers for the systematic review or because we were continuing to retrieve an overwhelming proportion of irrelevant material. We also wished to develop an estimate of the likely number of COS in the literature. Therefore, we examined the potential relevance of 1% of the hits (n=157). A random number generator (R) was used to select records, and their titles and abstracts were read to identify potentially relevant studies.

2.3.3 Electronic databases

This high yield of eligible studies from MEDLINE and evidence that no single database is likely to be sufficient for identifying research across health care [70-72], led to further work to select the sources to be searched. A variety of electronic databases were considered for searching:

1. MEDLINE via Ovid
2. The Cochrane Library (excluding categories 'Trials' and 'Cochrane Groups')
3. Cumulative Index of Nursing and Allied Health Literature (CINAHL) plus
4. Scopus

Medline via OVID

MEDLINE is the U.S. National Library of Medicine's (NLM) premier bibliographic database that contains over 19 million references to journal articles in life sciences with a concentration on biomedicine. Time coverage is generally 1946 to the present, with some older material. Currently it offers citations from approximately 5,600 worldwide journals (4,800 current biomedical journals) in 39 languages. The subject scope of MEDLINE is biomedicine and health, broadly defined to encompass those areas of the life sciences, behavioural sciences, chemical sciences, and bioengineering needed by health professionals and others engaged in basic research and clinical care, public health, health policy development, or related educational activities. MEDLINE also covers life sciences vital to biomedical practitioners, researchers, and educators, including aspects of biology, environmental science, marine biology, plant and animal science as well as biophysics and chemistry. Increased coverage of life sciences began in 2000.

The Cochrane Methodology Register

The Cochrane Library is a collection of six databases that contain different types of research: Cochrane Database of Systematic Reviews; Cochrane Central Register of Controlled Trials; Cochrane Methodology Register; Database of Abstracts of Reviews of Effects; Heath Technology Assessment Database, and NHS Economic Evaluation Database. Each of these has a different focus, with the Cochrane Methodology Register being most relevant to this project, but it should be noted that work on the development and maintenance of the Register was suspended by the UK Cochrane Centre in May 2012 and it has not been updated since July 2012. The Cochrane Methodology Register (CMR) is a database of studies relevant to the methods of systematic reviews of healthcare and social interventions. The register includes journal articles, book chapters, conference proceedings, conference abstracts and reports of ongoing methodological research. Relevant records are identified primarily through a programme of hand searching undertaken by the UK Cochrane Centre. The register aims to include all published reports of empirical methodological studies that could be relevant for inclusion in a Cochrane methodology review, along with comparative and descriptive studies relevant to the conduct of systematic reviews of healthcare interventions. In Issue 3, 2011 of *The Cochrane Library*, CMR contains 14,761 records.

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

This is the world's most comprehensive nursing & allied health research database, indexing for more than 5,000 journals. Offering complete coverage of English-language nursing journals and publications from the National League for Nursing and the American Nurses' Association, CINAHL covers nursing, biomedicine, health sciences librarianship, alternative/complementary medicine, consumer health and 17 allied health disciplines. Full-text coverage dates back to 1937.

Scopus

Scopus, launched in November 2004, is the largest abstract and citation database of peer-reviewed research literature. With over 19,000 titles from more than 5,000 international publishers (18,500 peer-reviewed journals including 1,800 Open Access journals), SciVerse Scopus offers researchers a quick, easy and comprehensive resource to support their research needs in the scientific, technical, medical and social sciences fields and, more

recently, also in the arts and humanities. Scopus includes all Embase and PubMed Journals from 1996 onwards.

Deciding which databases to search

MEDLINE focuses on biomedical journal literature, CINAHL on nursing and allied health literature, and Scopus covers medical and scientific literature, so we thought each had a sufficiently different focus to consider including in the search. Previous work that has been done to consider the coverage between CINAHL and Scopus to determine whether Scopus alone provides sufficient coverage of the literature concluded that only partial duplicate coverage of nursing and allied health literature was offered by CINAHL [70]. While SCOPUS's significantly larger coverage may offer many unique titles in these subject areas, it is not possible to say that these titles would be an adequate substitute for CINAHL's coverage of this literature. As its relevance to COS work was not yet known, it warranted further exploration. EMBASE was also considered for inclusion. We decided to include SCOPUS as opposed to EMBASE as it is a larger database and offers more coverage of scientific, technical, medical and social science literature[35]. Furthermore, SCOPUS indexes all EMBASE journals. The relevant modified search strategy was applied to each of these databases and each was considered in turn for suitability for inclusion in the final strategy.

2.3.4 Hand searching

In addition to the electronic database searching, we decided to complete a range of hand searching activities, in keeping with research evidence showing the added benefits of hand searching alongside electronic searching [73]. We identified and reviewed funded projects that included the development of a COS, including National Institute for Health Research (NIHR) programme grant scheme reports and Health Technology Assessment (HTA) reports; searched for known key authors and citations to key papers, for example, the work of the OMERACT group; examined references cited in eligible studies and in other studies that referred to or used a COS. We also contacted the Cochrane Review Groups (CRGs) to request information on COS that they were aware of (described in more detail in Chapter 3).

2.3.5 Sensitivity, precision and numbers needed to read (NNR)

We recorded whether each included study was:

i. retrieved by the search strategy developed for each database
ii. indexed on each database (regardless of whether or not it was retrieved by the search of the database)

The sensitivity (or recall), precision and numbers needed to read (NNR) for the final searches in each of the databases were calculated using the following definitions [74]:

Sensitivity (%) = 100 x (Number of included records retrieved/ Total number of included records)

Precision (%) = 100 x (Number of included records retrieved/ Total number of records retrieved)

NNR = 1/precision

Unique yield = number of studies retrieved only by this database

In addition, sensitivity*precision was calculated to allow a balance between sensitivity and precision to be assessed.

2.4 Results

The results of the search strategy development process are shown in Figure 2. When one author examined 1% (n=157) of the large number of retrieved records (n=15,704), 30 (19%) were identified as being potentially relevant, including three that we had previously identified as eligible studies. It was determined that 8 of the other 27 records were eligible following an assessment of their full papers. This confirmed that the search strategy was identifying relevant studies, that the likely yield of such studies was likely to be high and that a formal, systematic review was necessary to identify papers that were not yet known to COMET if we were to create a comprehensive resource that others could use to

determine whether or not a COS had already been developed in an area of interest to them.

The search strategy was modified as appropriate for each database. The combined results (searched 29 May 2012) generated a total of 47,225 records (MEDLINE n=14,520, Cochrane Library n=4122, CINAHL n=16,700, Scopus n=11,883), which fell to 37,132 after removal of duplicates (duplicates accounted for approximately 22% of total). We therefore needed to consider each of the additional databases that we were planning to search more carefully to estimate their likely added yield over MEDLINE. For example, the Cochrane Methodology Register includes articles that are relevant to the methods for systematic reviews, trials and other evaluations of health and social care and, as such, would be the most relevant component of The Cochrane Library for reports on the development of COS. With this in mind, the search strategy that was developed for The Cochrane Library as a whole was then limited to the Cochrane Methodology Register. It is recognised that The Cochrane Methodology Register has a unique controlled vocabulary but a decision was made to use the generic approach typically used when searching the whole Cochrane Library as this had already been developed for this search. For Scopus and CINAHL, two of the authors (EG and PW) independently reviewed a sample of abstracts from each. Fifty abstracts from Scopus yielded three eligible records that were not retrieved by MEDLINE or CINAHL. For CINAHL, a review of 100 abstracts excluded 91 as ineligible based on the abstract and the remaining nine potentially eligible studies were all identified by MEDLINE, Scopus or both. As a consequence, it was agreed that CINAHL would not be used, at least in this first round for the systematic review. Therefore, the following electronic databases were searched (August 2013):

1. MEDLINE via Ovid
2. SCOPUS
3. Cochrane Methodology Register

This updated search identified 34,398 potentially relevant records, all of which were checked and 220 eligible records were found (the process for selecting studies for inclusion in the review is fully described in Chapter 3). Fifty-nine (27%) were already known to us, so the search identified an additional 161 records. In addition to the database search, 30 additional records that had not been previously identified were deemed eligible after being identified through hand searching. A full list of the 250 included records is provided in

Appendix 2. Sensitivity of the search strategies ranged from 4% to 86%, and precision from 0.8% to 1.1% (Table 1). MEDLINE via Ovid performed best in terms of sensitivity, retrieving 216 (86%) of the 250 included records, followed by Scopus (44%). The search of the CMR identified just 4% of the included records, and all of these were found in at least one of the other databases. MEDLINE via Ovid was also the database with the highest precision. The number needed to read varied between 89 (MEDLINE via OVID) and 130 (SCOPUS). If our searches had been limited to MEDLINE alone, only 3 included records unique to SCOPUS would not have been retrieved.

Table 1: Sensitivity, precision and NNR for each strategy

Database	Number of records retrieved	Number of included records	Unique yield^	Sensitivity (%) (n=250)	Precision (%)	Number Needed to Read (NNR)	Sensitivity *Precision
MEDLINE via OVID	19058	216	109	86	1.1	89	98.3
SCOPUS	14258	109	3	44	0.8	130	33.9
CMR	1082	9	0	4	0.8	121	3.3

^ *Unique yield relates to the number of studies retrieved by one database only.*

Through examining references cited in eligible studies and in other studies that referred to or used a COS, 30 additional records were identified and included in the systematic review. No additional studies were identified through the survey of Cochrane Review Groups. Of the 30 records identified via hand searching and not retrieved by the search strategy, we found that two were not indexed on any database, 25 were indexed on both Medline and Scopus, two on Scopus only, and one was in all three databases. On closer inspection, we found that the reasons for non-retrieval of the 28 studies by the database searches was the wide variety of free text and index terms used in their records. As has been shown in other contexts, modifications to the search to retrieve all these records would have produced searches with unmanageably large numbers of records [75]. Furthermore, the absence of two of the reports from the searched databases highlights that even such extensive searching would not have retrieved all the studies that we identified. However, one of these was not a journal article, and the other was an editorial which provided additional methodological information on a study that had been retrieved in its own right.

2.5 Discussion

COS are increasingly recognised as important for the design, conduct and reporting of randomised trials, systematic reviews and other forms of research. However, as I have shown in this chapter, the development of a search strategy to identify them is challenging. A search for COS in any specific area could combine the approach we have taken with search terms for specific conditions or interventions, but it is still likely to require a large number of records to be checked to identify the few that are eligible. We hope therefore that this comprehensive approach to searching a major database such as MEDLINE for all reports of studies developing COS, regardless of the setting, and the subsequent inclusion of identified studies in the COMET database will make it much easier for researchers in the future. This is akin to the work of The Cochrane Collaboration in identifying reports of randomised trials regardless of topic area for inclusion in the Cochrane Central Register of Controlled Trials [76, 77].

In undertaking this comprehensive approach to identifying COS, similar challenges were encountered to those faced by healthcare researchers in the past. For example, variability in the use of free text terms and index terms on reports of randomised trials of portal vein infusion chemotherapy in colorectal cancer meant that a search to identify all articles that had been identified for a systematic review would have had to rely solely on the terms for colorectal cancer, which retrieved 18,450 records [75]. On a larger scale, when The Cochrane Collaboration was established in 1993, although tens of thousands of reports of randomised trials could be found easily in MEDLINE, there were many more than had not been appropriately indexed and could not be found so easily. The development of highly sensitive search strategies and subsequent work within the Collaboration to find these "hidden" reports, led to the identification of an additional 70,000 records that were re-tagged as randomised controlled trials or controlled clinical trials in MEDLINE and can now be found using those terms [76, 77]. In a similar way, this comprehensive searching for COS regardless of any particular healthcare condition and their inclusion in the COMET database makes it much easier for users to access these studies.

Although work on the development of COS goes back at least 30 years [41], the term itself has not been widely used until relatively recently and there are currently no MeSH headings in MEDLINE or index terms in other bibliographic databases for identifying COS papers, and they do not appear to be categorised consistently across different databases.

Furthermore, no single database specialises in this type of methodological research and it is likely to be found across a wide range of literature. For example, MEDLINE and EMBASE focus on biomedical journal literature, CINAHL on nursing and allied health literature, and Scopus covers medical and scientific literature. Each database has a different focus but each could include studies of the development of COS. Furthermore, no single database is likely to be adequate. For instance, a comparison of the coverage between CINAHL and Scopus to determine whether Scopus alone provides sufficient coverage of the literature found that Scopus can only partially duplicate the coverage of nursing and allied health literature offered by CINAHL [70]. The work in this chapter did not identify any unique yield in the sample checked for CINAHL, so it was agreed that CINAHL would not be used, at least in this first round for the systematic review. However, due to the reasons listed here, further appraisal of CINAHL could be considered in any future updates of this search.

Other comparisons, of other combinations of databases, have also shown how systematic reviews are likely to benefit from searching for potentially eligible studies in several databases [71, 72].

The search strategy developed for COS has been designed to be highly sensitive, so that as many potentially relevant studies as possible will be retrieved. The final effective search strategy combines keywords, index terms and free-text terms and phrases, using combinations of Boolean operators. As no MeSH headings or index terms currently exist for COS papers and these papers do not appear to be indexed in a consistent way, key search terms were also targeted and no limits were applied to the search. A consequence of a highly sensitive search is usually that a large number of irrelevant records will be retrieved, the majority of which will likely not meet the inclusion criteria for the review, and this appears to be the case here.

I found that two databases and hand searching were required to locate all of the studies that were included in this review. MEDLINE via Ovid alone retrieved 86% of the included studies, but actually 97% of the included studies were indexed on MEDLINE. The search of the Cochrane Methodology Register did not identify any records that were not found in the other databases. I identified retrospectively that the search in The Cochrane Library used an implied ‘AND’ in some of the lines which could affect the sensitivity and precision of this search. Given that its development and maintenance was suspended in July 2012, it will not

be included in future searches to identify studies developing COS. However, this decision will be reviewed and the search strategy evaluated should work on the Register resume. SCOPUS had the lowest precision rate (0.8) and highest number needed to read (130), which is a particularly high number of records to check in order to find one relevant record. Therefore, with such a low unique yield from SCOPUS (3 studies identified with the search strategy, increasing to 5 actually indexed); it might not be worth searching SCOPUS, given the low precision and high NNR. Furthermore, the decision to search SCOPUS may have increased the de-duplication burden due to the lack of flexibility in its interface, for example de-duplicating against MEDLINE within OVID. This could be an added advantage of searching EMBASE and may be considered in any future updates to this search.

2.6 Conclusions

In considering how this analysis might inform future decisions about the COMET searches for COS, a balance needs to be struck between the work involved in screening large numbers of records, the frequency of the searching and the likelihood that eligible studies will be identified by means other than the database searches. One possibility is that the comprehensive searching is limited to MEDLINE, perhaps with further appraisal of the unretrieved records to seek ways to modify the search to target these papers. The search might also continue to include SCOPUS, but consideration could be given to other databases that were not included in this project, such as EMBASE. In keeping with the research evidence, it would seem there is an added benefit to hand searching, so this should accompany any electronic searching in the future. This would help to identify relevant studies that are missed by the database searching, either because they are not indexed in the databases or are indexed but cannot be retrieved without using a search strategy that would yield an unmanageable number of records to check. To keep the database current, COMET aims to complete an annual search of MEDLINE and SCOPUS as a minimum. Finally, to supplement this annual search, the COMET database will continue to be populated with studies, both ongoing and completed, that are identified by ad hoc means and sent directly to the COMET team.

The full methods and results of this systematic review are presented and discussed in Chapter 3.

Chapter 3: A systematic review of core outcome sets

3.1 Background

For COS to be an effective solution to problems of inconsistency and outcome reporting bias in trials, they need to be accessible to researchers and other key groups. COS are currently scattered across the health literature, which means the identification of existing COS is not easy. This chapter describes a systematic review to attempt to bring together existing COS together for the first time. The development of the search strategy was described in Chapter 2. This chapter provides a detailed description of the methods used and the results of the systematic review.

3.2 Aims

To identify studies which had the aim of determining which outcomes or domains to measure in all clinical trials in a specific condition, and to identify and describe the methodological techniques used in these studies.

3.3 Methods of the review

The protocol is available at http://www.comet-initiative.org/about/researchprojects.

3.3.1 Inclusion and exclusion criteria

Studies were eligible for inclusion in the review if they had developed or applied methodology for determining which outcome domains or outcomes should be measured, or are important to measure, in clinical trials or other forms of health research.

Studies were ineligible if they did not do this but were related to how, rather than which, outcomes should be measured; reported the design or rationale for a single trial or for preclinical or early phase trials only; reported the use of a core outcome set*; were a systematic review of clinical trials; were studies or systematic reviews of studies of

prognosis; were studies (including systematic reviews and surveys) of outcomes measured in clinical trials** or quantitative descriptions (e.g. frequency) of outcomes**; were based on the opinion of a single author only**or focussed on one domain/outcome only**.

** reports relating to COS but not meeting inclusion criteria (e.g. where a COS had been used) were retrieved, and their references checked for potentially eligible studies.*

*** although these were not included in the systematic review, they were eligible for inclusion in the COMET database.*

3.3.2 Types of participants and interventions

Studies relating to participants of any age, with any health condition in any setting and assessing the effect of any intervention were eligible for inclusion.

3.3.3 Identification of relevant studies

In order to optimise the identification of relevant studies, the following databases were searched in August 2013:

1. MEDLINE via Ovid
2. SCOPUS
3. Cochrane Methodology Register

We developed a multi-faceted search strategy to search these databases using a combination of text words and index terms, adapting the search strategy as appropriate for each database. The process of developing an appropriate search strategy was described in Chapter 2. For full details of the search strategy see Appendix 1.

In addition to the electronic database searching, we completed a range of hand searching activities, in keeping with research evidence showing the benefits of adding hand searching to electronic searching [73]. We identified and reviewed funded projects that included the development of a COS, including NIHR programme grant scheme reports and HTA reports; searched for known key authors and citations to key papers, for example, the work of the OMERACT (Outcome Measures in Rheumatology) group; examined references cited in

eligible studies, and examined references cited in other reports relating to COS that were not eligible for this review (e.g. where a COS had been used or referred to).

We contacted the 50 Cochrane Review Groups (CRG) as of 2011 across all areas of health care to request information on COS that they are aware of (by asking "Are you aware of any other work already done/being done attempting to develop a core outcome set for conditions covered by your CRG?"). Full details of the methods used for that study can be found in Kirkham et al 2013 [32].

3.3.4 Selecting studies for inclusion in the review

The records retrieved from searching were combined, and duplicate records removed. Their titles and abstracts were then read to assess eligibility (stage 1) and the full texts of potentially relevant articles were obtained and assessed for inclusion (stage 2).

Stage 1a

The title and abstract of each citation was initially read by one reviewer (EG). Each citation was categorised as include, unsure, exclude, exclude-references to be checked, or exclude-suitable for COMET database.

Stage 1b

Those categorised as include or unsure were then checked independently by a second reviewer (BG) and categorised as above. The assessments of the two reviewers were compared, and the records were categorised again following discussion. If agreement could not be achieved, the citation was categorised as unsure and retained for future checking. Where there was agreement to exclude, the citation was excluded at this stage. Full papers were retrieved for all records categorised as include, unsure, or references to be checked after this screening of the title and abstract.

Stage 2

Each full paper was checked by one of three reviewers (EG, BG, or NM) and categorised as above. Citations categorised as unsure were checked by a second reviewer and discussion took place between the reviewers to reach consensus on eligibility. An additional reviewer was consulted if agreement could not be achieved (PW). Reasons for exclusion were documented for each article judged to be ineligible in stage 2.

Full papers were obtained for all records categorised as relating to COS based on the title and abstract but which do not appear to meet the inclusion criteria (e.g. where a core outcome set had been used in a research study). Reference lists were checked, and relevant citations retrieved and assessed for inclusion. The reference lists in the reports of all included studies were also checked, and relevant citations were retrieved and assessed for inclusion.

3.3.5 Checking for agreement between reviewers

At stage 1 (stage 1a done by EG for all records and using all the categories, and stage 1b when BG checked records that had been categorised as include or unsure), reviewers independently assessed batches of 10 abstracts (EG and BG) to check for agreement before independently assessing records. Results were then compared and discussed. This was repeated in batches of 10 until complete agreement was reached. Following this, a further 50 abstracts were reviewed and compared to confirm agreement between the two reviewers, who then went on to assess all abstracts that had been initially recorded as include or unsure.

At stage 2, reviewers independently assessed batches of 10 full papers (EG, BG, NM). These were then compared and discussed with the lead reviewer (EG). This was repeated until full agreement was achieved in three consecutive batches. Reviewers then assessed papers individually.

3.3.6 Checking for correct exclusion

Full papers were obtained for a 1% sample of the records excluded on the basis of the title and abstract (stage 1) and checked for correct exclusion by a second reviewer (NM). If any studies were found to have been excluded incorrectly, additional checking would be performed within the other excluded records. A reviewer also assessed a minimum of 5% of the papers that were excluded after reading their full text, to check for correct exclusion at stage 2.

3.3.7 Data collection and extraction

A COS may be developed to cover all aspects of a disease or health condition, but it may also have been developed with a focus on a particular type of treatment only, or for a

specific age group or stage of disease. It is therefore important in reporting the scope of a COS to consider the specific area of health or healthcare to which it applies, along with details of health condition, population (here we focussed on age) and types of interventions [38]. We therefore extracted the following data as free text unless otherwise stated:

1. Study details, including
    a. Year of publication
    b. Study aims
    c. Intended use of recommendations

2. Health area
    a. Disease or health category (e.g. 'Lungs & airways' or 'Pregnancy & childbirth') [using a checklist]
    b. Disease name (e.g. 'Asthma')

3. Target population
    a. Age
    b. Type of intervention

4. Method of development used
    a. Rationale for choice of method(s) used

5. Definition of consensus
    a. Rationale for definition used

6. Rating/scoring system used
    a. Rationale for rating/scoring system used

7. Stakeholders involved in the process (e.g. health professionals, patients*, industry)
    a. Geographical setting of participants

8. Outcomes included in the core set

9. Plans for implementation

10. Plans to review or update the core outcome set

11. Was how to measure outcomes considered?

** In using the term 'patient' throughout this thesis, I include patients, carers, health and social care service users and people from organisations who represent these groups. I selected the term 'patient' in preference to alternatives such as 'public' [78] as patient is more in keeping with the terminology used by COS developers, and with the capacity with which this group of participants were involved in COS development.*

3.3.8 Data analysis and presentation of results

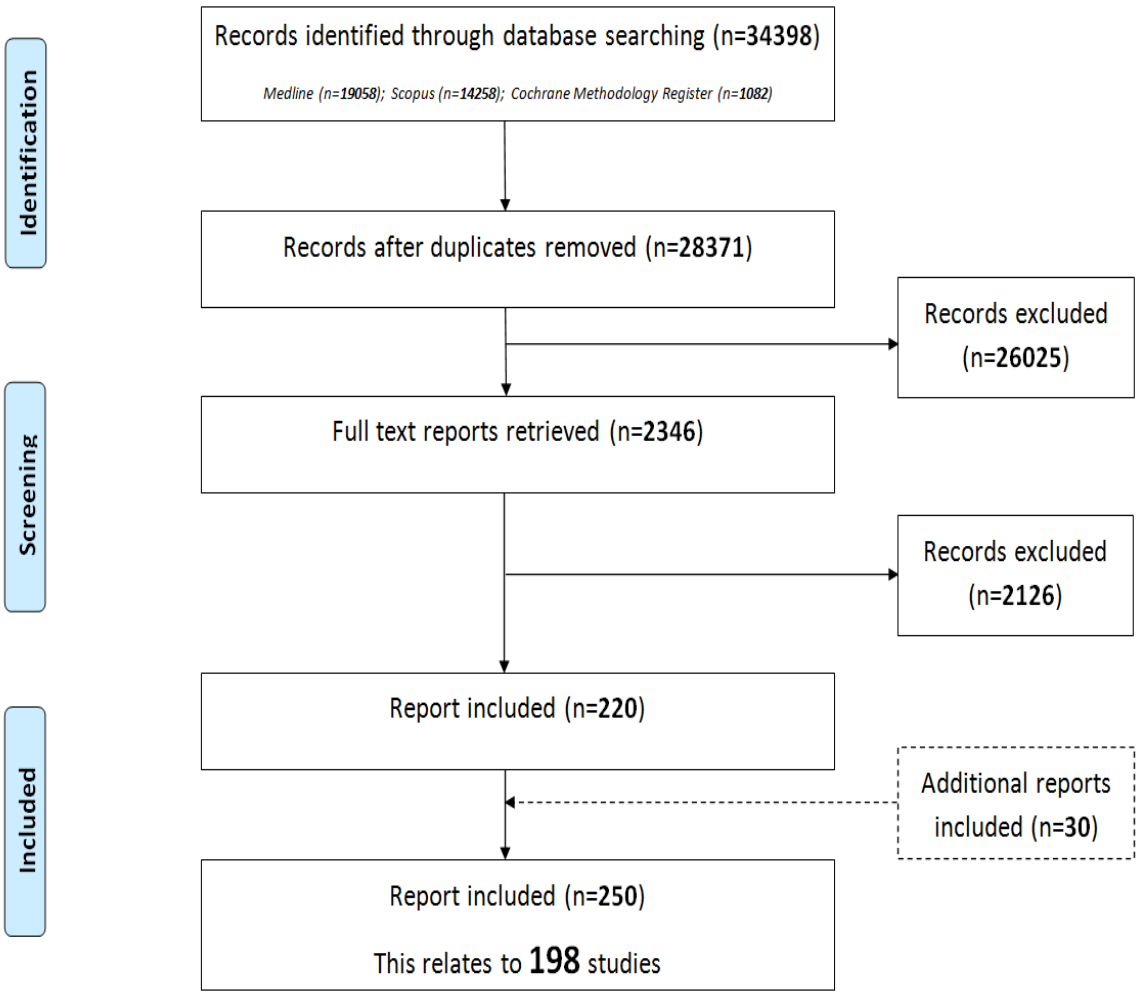
Studies were described in narrative form, and the findings provided in text and tables format. We did not anticipate conducting any statistical analyses to combine the findings. Reporting was carried out in accordance with PRISMA guidelines [79]. PRISMA stands for Preferred Reporting Items for Systematic Reviews and Meta-Analyses. It is an evidence-based minimum set of items for reporting in systematic reviews and meta-analyses.

3.4 Results

3.4.1 Description of studies

The initial database search identified 28,371 potentially relevant records after duplicates had been removed. Screening at stage 1 and 2 excluded 26,025 and 2126 records respectively. The process of identifying studies is summarised in Figure 3. This large number of excluded records was expected because the search was intentionally over-inclusive due to the lack of index terms for such studies and the likelihood that many records would be retrieved merely because they discussed issues around the use of outcomes in trials, but without the necessary emphasis on how these outcomes were selected. A summary of the reasons for exclusion at stage 2 is presented in Table 2. Two-hundred and twenty citations met the inclusion criteria. In addition to the database search, 30 additional citations were deemed eligible after being identified through reference checking. No additional studies were identified through the survey of Cochrane Review Groups. In total, 250 reports relating to 198 studies were included in the review.

Figure 3: Identification of studies



The relevant tables and figures in this chapter are reproduced with permission from Gargon, E., B. Gurung, et al. (2014). "Choosing important health outcomes for comparative effectiveness research: a systematic review." PLoS ONE 9(6): e99111.

Table 2: Reasons for exclusion at stage 2 (assessment of full text reports)

Reason	N
Review/overview/discussion only, no outcome recommendations	495
Core outcomes/outcome recommendations not made	214
HRQL*[1]	117
How to measure outcome (including instruments, tools, scales, scores, outcome definition)	123
ICF core set development*[2]	80
Quality indicators – included an aspect of outcomes*[3]	78
Not relevant	669*[5]
ICF core set validation	56
Quality indicators – structure and/or process of care only	52
One outcome/domain only	40
Clinical management in practice not research (including for diagnosis)	45
Instrument development	24
Recommendation by single author only	21
Registry development*[4]	21
Describes features of registry	16
Preclinical/ Early phase only (0, I, II)	18
On-going work	11
Duplicate	11
Quantitative description (e.g. frequency of symptoms)	9
Reporting the design/rationale of a single trial	22
Oral presentation only	2
Value attributed to outcomes	2
TOTAL	**2126**

**Although these studies are relevant to the development of a COS (and therefore suitable for inclusion in the COMET database), they did not meet the review inclusion criteria.*

**[1] These studies included qualitative studies describing the impact of a treatment or condition on a patient's quality of life, studies to determine particular domains of quality of life, and single patient narratives of the impact of a condition or treatment on their quality of life. The focus of these studies was on quality of life only.*

**[2] Although the ICF is widely comprehensive, it is not all inclusive. For example, the ICF does not include outcomes such as death, an outcome that is often relevant to measure in clinical trials. Furthermore, as the ICF focuses on the individual only, caregiver outcomes would not be included. While for many health areas this may not be relevant, for some (e.g. dementia), caregiver outcomes may be core to measure. See inclusion and exclusion criteria in Section 3.3.1, and further discussion of ICF core set studies in Section 3.4.4.*

**[3] These studies assessed quality or efficiency of care (clinical practice), or the performance of an individual institution. Indicators were often specific to that scenario/environment of care only.*

**[4] These studies described the development of registries, each with its own purpose, often to evaluate management of patients, identify best practices or to describe therapeutic strategies.*

**[5] 599 of these (90%) had no abstract to assess (title only), so had to be reviewed at full paper due to potential eligibility based on the title alone.*

3.4.2 Agreement between reviewers

At stage 1, reviewers independently assessed batches of 10 abstracts (EG and BG), and the results were then compared and discussed. This was repeated in batches of 10 until complete agreement was reached. Five batches of ten were reviewed before reaching full agreement, with agreement of 70%, 80%, 80%, 80%; and finally 100%. Following this, a further 50 abstracts were reviewed and compared to confirm complete agreement between reviewers. Of the 2238 abstracts reviewed at this stage, the two reviewers agreed 98.7% of the assessments. The reviewers disagreed on 29 abstracts, so the full text for these were obtained and reviewed.

At stage 2, reviewers independently assessed batches of 10 full papers (EG, BG, NM). These were then compared and discussed with the lead reviewer (EG). This was repeated until full agreement was achieved in three consecutive batches. The lead reviewer had full agreement in the first three batches with one reviewer and in the ninth, tenth and eleventh batches with the other.

3.4.3 Checking for correct exclusion

308 of the 26,025 (1.2%) abstracts excluded at stage 1 were reviewed by a second reviewer to check for correct exclusion from the review, and assessed for inclusion in the COMET database. None were found to have been excluded incorrectly, and no further checks were performed. 115 of the 2127 (5.4%) full papers excluded at stage 2 were reviewed and again assessed for inclusion in the COMET database. One study was found to be incorrectly excluded. Given this low error rate (less than 1%), it was agreed that it would not be efficient to undertake any further checking of the excluded full papers.

3.4.4 Excluded studies

Studies outside the scope of this review which were found in the searches include International classification of Functioning, Disability and Health (ICF) core set development papers, quality indicator studies that included an aspect of outcomes, and registry development. Although these studies were not eligible for inclusion in the review they might be relevant to COS development and they will be included in the COMET database. Of particular interest were the ICF development papers. ICF offers a framework to describe functioning, disability and health in a range of diseases. The ICF focuses on the assessment

of an individual's functioning in day-to-day life. It provides a framework for body functions, activity levels and participation levels in basic areas and roles of social life; providing domains of biological, psychological, social and environmental aspects of functioning [23]. In many clinical areas, ICF core sets have been developed. These core sets identify the most relevant ICF domains for a particular health condition. We considered the inclusion of these core sets in this systematic review but we decided not to do so (but they will be added to the COMET database). Although the ICF is widely comprehensive, it is not all inclusive. For example, the ICF does not include outcomes such as death, an outcome that is often relevant to measure in clinical trials. Furthermore, as the ICF focuses on the individual only, caregiver outcomes would not be included. While for many clinical areas this may not be relevant, for some (e.g. dementia), caregiver outcomes may be core to measure. For these reasons, ICF core sets have not been included in the systematic review, but I provide an overview of the clinical areas where an ICF core set has been developed (Table 3).

Table 3: ICF core sets

Long-term context
Chronic widespread pain
Low back pain
Osteoarthritis
Osteoporosis
Rheumatoid arthritis
Chronic ischemic heart disease
Diabetes Mellitus
Obesity
Obstructive pulmonary diseases (COPD)
Depression
Breast cancer
Stroke
Ankylosing spondylitis (AS)
Spinal cord injury
Systemic lupus erythematosus (SLE)
Multiple sclerosis
Head and neck cancer
Sleep
Hand conditions
Bipolar disorders
Traumatic brain injury
Inflammatory bowel disease
Amputation
Hearing loss
Vertigo, dizziness and balance
Cerebral Palsy

Situation-specific context
Vocational rehabilitation
Acute context
Neurological conditions
Musculoskeletal conditions
Acute inflammatory arthritis
Spinal cord injury
Cardiopulmonary conditions
Geriatric patients

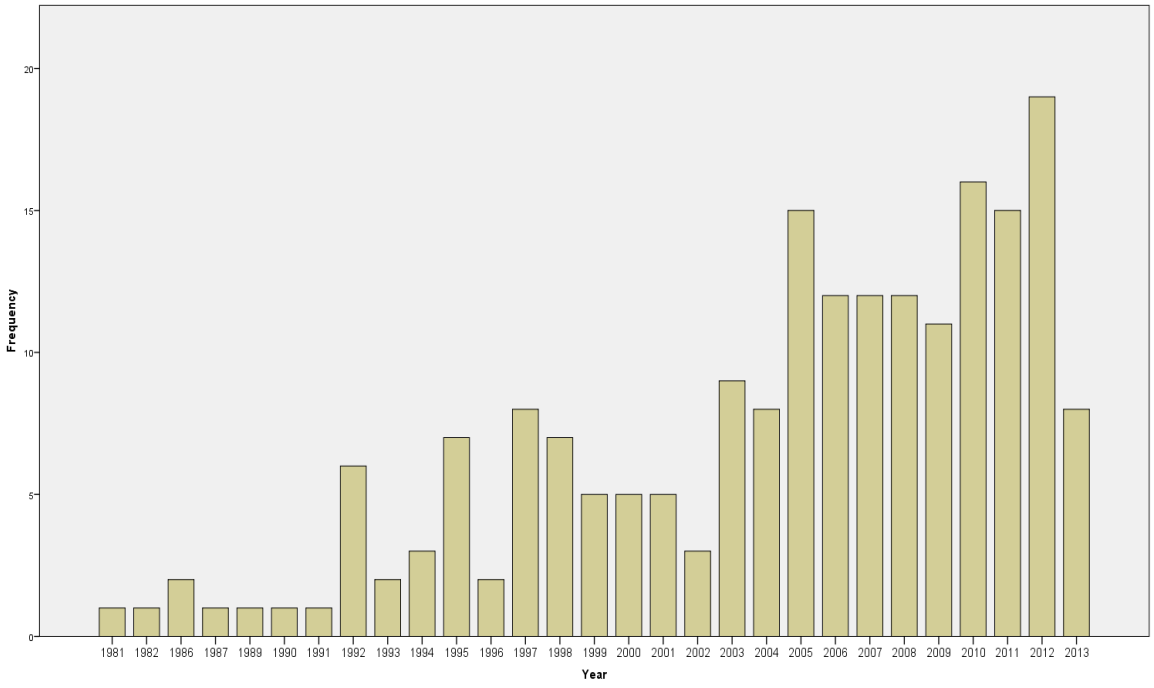
3.4.5 Characteristics of included studies

A total of 198 studies (250 reports) were included (Appendix 2). The number of reports per study ranged from 1 to 8. Most studies had a single report only (168/198; 85%) and eighteen studies had two corresponding reports. Three studies describing the work carried out on behalf of the Neonatal Drug Development Initiative (NDDI) were all linked with the same additional report that further described the methods. Three studies had three corresponding reports, three studies had four corresponding reports, and there were six, seven and eight reports for single studies. Reports were typically published in disease specific journals; the journal of publication is listed in Appendix 2.

3.4.6 Year of publication

The year of publication of the earliest identified report for each study is shown in Figure 4, which clearly shows a general increase in the number of COS over the years.

Figure 4: Year of first publication of each study (N = 198)



2013 Includes January to June only.

3.4.7 Scope of core outcome sets

A COS may be developed to cover all aspects of a disease or health condition, but it may also have been developed with a focus on a particular type of treatment only, or for a specific age group or stage of disease. It is therefore important in reporting the scope of a COS to consider the specific area of health or healthcare to which it applies, along with details of health condition, population (age) and types of interventions [38]. The scope of included studies is summarised in Table 4. This includes study aims, intended use, disease categories (classification according to disease name can be found in Appendix 2), population characteristics and intervention characteristics.

Half of the studies considered outcomes whilst addressing wider clinical trial design decisions (51%), and half specifically considered outcome selection and measurement (49%). Most studies developed recommendations intended specifically for clinical trials (141/198; 71%); however a further 27 studies (14%) intended their recommendations for health research more generally. Some studies intended their recommendations for clinical practice as well as trials or research generally (n=10 and n=11, respectively) and for regulatory purposes (n=3). Notably, one study explicitly stated they intended their recommendations for use in all of these situations.

With regard to the age of the population that a COS was intended for, the majority of studies (149/198; 75%) did not specify whether their recommendations were specific to children, adults or both. Thirteen (7%) made an explicit statement that they intended their recommendations for both children and adults. Some studies made recommendations explicitly for one population or the other. Twenty-three studies (12%) specified they related to children only, and 13 (7%) to adults only, with three of these specific to 'older adults'.

Again, when we consider the intervention scope, the majority (115/198; 58%) did not specify their intentions. Only seven studies (4%) made recommendations that were clearly intended for all intervention types. The remaining studies (38%) specified the intervention type to which the COS should apply, including drugs (n=40), surgery (n=13) and vaccines (n=2).

Table 4: The Scope of included studies (N=198)

	N (%)
Study aims	
Considered outcomes while addressing wider clinical trial design issues (e.g. trial duration, ethical issues, eligibility criteria etc.)	101 (51)
Specifically addressed outcome selection and measurement	97 (49)
Intended use of recommendations	
Clinical trials	141 (71)
Clinical research	27 (14)
Clinical research and practice	11(6)
clinical trials and clinical practice	10 (5)
Clinical trials and regulatory purposes	3 (2)
Trials and observational studies	3 (2)
Observational studies	1 (<1)
Trials and case series	1 (<1)
Clinical research, clinical practice and regulatory purpose	1 (<1)
Disease Categories	
Cancer	31 (16)
Rheumatology	28 (14)
Neurology	24 (12)
Heart & circulation	22 (11)
Dentistry & oral health	12 (6)
Infectious disease	12 (6)
Orthopaedics & trauma	10 (5)
Lungs & airways	8 (4)
Gastroenterology	8 (4)
Gynaecology	6 (3)
Tobacco, drugs, & alcohol dependence	4 (2)
Urology	4 (2)
Blood disorder	3 (2)
Anaesthesia & pain control	3 (2)
Mental health	3 (2)
Neonatal care	3 (2)
Skin	3 (2)
Others (chronic conditions, benign disease, intensive care)	3 (2)
Kidney disease	3 (2)
Pregnancy & childbirth	2 (1)
Endocrine & metabolic	2 (1)
Ear, Nose & Throat	1 (<1)
Genetic disorders	1 (<1)
Wounds	1 (<1)
Health care of older people	1 (<1)
Population characteristics	
All (adults and children stated explicitly)	13 (7)
Children	23 (12)
Adults	10 (5)
Older adults	3 (2)
Not specified	149 (75)
Intervention characteristics	

All intervention types	7 (4)
Drug treatments	40 (20)
- *Drug only*	*34*
- *Drug, and rehabilitation*	*1*
- *Drug and delivery management*	*1*
- *Drug and physical therapy*	*1*
- *Drug and complementary and alternative medicine (CAM) treatment*	*1*
- *Immunomodulatory therapies*	*2*
Vaccine	2 (1)
Surgery	13 (7)
Procedure*	5 (3)
Device**	3 (2)
Other***	13 (7)
Not specified	115 (58)

Procedure descriptions –
Procedure - Uterine artery embolization
Procedure - Aortic valve stenosis (AS) - transcatheter aortic valve implantation
Procedure - Aortic valve stenosis (AS)
Procedure - pulp treatments of primary teeth
Procedure - drug-eluting coronary stents (DES)

Device descriptions –
Device – Compression (n=2)
Device - Mechanical circulatory support (MCS)

****Other descriptions –***
Coronary angiogenesis
Hip protectors
Neuro-protective therapy (aka Neuroprotection)
Non-surgical treatment (no other detail given)
Operative and non-operative management
Oral care products
Ascorbic acid
Exercise/physical activity
Fall injury prevention interventions
Behavioural therapies or other kinds of nonpharmacologic therapies
Psychological & behavioural: Psychosocial
Rehabilitation (vocational)
Maternity care

Studies were classified according to disease category and disease name, and the various disease categories for which COS have been recommended are shown in Table 4. The most common disease categories were cancer (n=31), rheumatology (n=28), neurology (n=24), heart and circulation (n=22), dentistry and oral health (n=12), and infectious disease (n=12); but most disease categories (63%) had less than five COS. Although the majority of diseases had only one COS (n=142), some diseases had more than one corresponding core set. These are summarised in Table 5.

Table 5: Number of COS per disease

Number of COS	Number of diseases	Same scope	Differences in scope	Detail of differences
1	142	0	142	
2	18	6	12	Differences in intervention (n=7), age (n=4) or setting (n=1).
3	2	1	1	One core set developed specifically for children, but age not specified for the other two core outcome sets.
4	2	1	1	One core set developed specifically for use in children while the other three sets did not appear to differ in their scope.
6	1	0	1	One developed specifically as a patient core set. The remaining five sets did not differ in their intended scope, but one of these was developed specifically to address this very issue.

3.4.8 Methods used to select outcomes

Studies reported using a variety of methods, sometimes in combination, to select the outcomes for the COS. There was no description of the methods used in 16/198 (8%) studies. The frequency of the different methods used to select outcomes in the included studies is provided in Table 6. The most frequent method used was semi-structured group discussion (n=104, 54%), which included workshops (n=39), meetings (n=60), and round table discussion (n=5). A further 23 studies were classified as using an unstructured group discussion (12%); descriptions included task forces, work(ing) groups/parties, committees, boards and panels. These studies did not describe whether they had face-to-face, telephone or electronic discussions. Sixty-five studies (33%) carried out a literature or systematic review. This was done in combination with another method in 54 of these 65 studies (83%). Other frequently used methods included the Delphi technique (n=29, 15%), Consensus Development Conference (n=20, 10%), surveys (n=17, 9%) and Nominal Group Technique (n=15, 8%). More than one method was used in 74/198 (37%) studies. More

detailed description about the combination of methods used can be found in Table 6. More detailed information about methods is provided in Chapter 4.

Table 6: The methods used to develop core outcome sets

Main methods	N (%)
Semi-structured group discussion only	**57 (29)**
- *Workshop*	*22*
- *Meeting (meeting, colloquium, conference where not described as consensus development conference)*	*32*
- *Round table discussion*	*3*
Unstructured group discussion only *Descriptions include task force, work group, working group/party, committee, board, panel*	**18 (9)**
Literature/systematic review only	**11 (6)**
Consensus development conference only	**12 (6)**
Delphi only	**6 (3)**
Survey only	**3 (2)**
NGT only	**1 (1)**
No methods described	**16 (8)**
Mixed methods – *see below*	**74 (37)**
Delphi + another method(s)	**23 (31)**
- *NGT*	*4*
- *NGT + Literature/systematic review*	*4*
- *Semi-structured discussion (meeting& Workshop)*	*2*
- *Systematic review + Survey*	*1*
- *Literature/systematic review*	*5*
- *Literature/systematic review + semi-structured group discussion (meeting/workshop)*	*3*
- *Literature/systematic review + Meeting(s) + focus group(s) + workshop*	*1*
- *Literature/systematic review + consensus conference*	*1*
- *Literature/systematic review + survey + meeting*	*1*
- *Meeting + survey*	*1*
Semi-structured group discussion (listed which method) + another method(s)	**29 (39)**
- *Workshop + literature/systematic review*	*4*
- *Meeting + literature/systematic review*	*13*
- *Workshop and meeting*	*2*
- *Workshop/meetings + web-based consultation*	*2*
- *Workshop, literature/systematic review*	*1*
- *Workshop + survey + literature/systematic review*	*3*
- *Round table discussion + literature/systematic review*	*2*
- *Meeting + focus group(s) + survey*	*1*
- *Meeting + survey*	*1*
Consensus development conference + another method(s)	**7 (10)**
- *Survey*	*1*
- *NGT*	*1*
- *Literature/systematic review*	*3*
- *Meeting(s)*	*1*
- *Literature/systematic review + survey + meeting*	*1*
Unstructured group discussion + Literature/systematic review	**5 (7)**
NGT + another method(s)	**5 (7)**

- *Survey + interview*	*1*
- *Semi-structured discussion (workshop & meetings)*	*1*
- *Survey*	*1*
- *Workshop + Literature/systematic review*	*1*
- *Literature review*	*1*
Survey + Literature/systematic review	**1 (1)**
Focus group + rating exercise	**1 (1)**
Literature/systematic review, public presentation and debate	**2 (3)**
Literature/systematic review, survey and open discussion	**1 (1)**

Furthermore, we found that of the 178 studies that described the methods they used to determine the COS, 164 (92%) did not provide an explanation regarding their choice of methodology. Rationale for methodological choice for the 14 studies that did provide explanation is provided in **Error! Not a valid bookmark self-reference.**. The most common reason provided was that the methods had been used previously to develop COS or were well-recognised methods for eliciting expert consensus to form guidelines.

Table 7: Rationale for methodological choice

Reference	Method(s) used	Efficient process	Minimise burden on participants	Higher response rate	Allow individual contribution	Minimising dominance by individuals	Larger sample	International participation	Interaction between participants	Allow multiple stakeholders	Previous work
Fried et al. (1993)	NGT	X			X				x	x	
Sinha et al. (2012)	Delphi										x
	Survey		x				x				
Devane et al. (2007)	Delphi	X		x				x			
Douglas et al. (2009)	Delphi, NGT										x
Khanna et al. (2008)	Delphi, NGT										x
Bellomo et al. (2004)	Consensus conference										x
Mease et al. (2005)	Delphi									x	x
Lux and Osborne (2004)	Delphi								x		x
Ruperto et al. (2003)	Delphi, NGT										x
Schmitt et al. (2011)	Delphi					x					
Moniz-Cook et al. (2008)	Web based consultation							x		x	
Cross (2005)	Delphi									x	
Smaïl-Faugeron et al. (2013)	Delphi										X
Distler et al. (2008).	Delphi							x			

3.4.8 Definition of consensus

Definition of consensus was provided in 19 studies; three studies explicitly stated that this was decided a priori. Agreement was typically defined as the proportion of participants in agreement, where the level of agreement ranged between 50% and 100% (see Table 8). In one study the level of agreement for the meeting part of the study was 'less than 30% of the voters disagree.' Although a definition of consensus was not specified in the remaining studies, there was some description about how the COS were formed in eight studies. In four studies this was described in terms of 'majority'; majority vote (n=2) majority opinion

(n=1) and majority agreement (n=1); one study simply defined agreement as ‘unanimous agreement’ and the final study by ‘vote’. Two studies described an ‘iterative process’ leading to consensus, but this was not defined.

Three studies provided a rationale for the consensus definition used. One attributed this to ‘common sense,’ one because the threshold had been used previously, and the third stated: ‘We chose 66% instead of the traditional 80% as there were only nine members in the Steering Committee and an 80% consensus would have required agreement among eight of nine experts. This latter was considered an unrealistic and overly stringent requirement.’

Ten of these nineteen studies used Likert scales to determine the importance or appropriateness of the outcomes (scales ranged from 5 to nine point scales) during the consensus process. Three studies used 100 points distribution; two studies used top three ranking and one study asked for top ten ranking of outcomes. No study provided a rationale for the choice of scale used. These 19 studies used different methods. The definition of consensus and rating scales used, specifically with the Delphi method and those that followed this with a consensus meeting, are discussed in more detail in Chapter 4.

Table 8: Definition of consensus and type of rating scale (N=19 studies)

Reference	Method	Use of rating scales	Consensus (value in %)	Rationale for consensus definition
Bellamy (1997)	Consensus development conference	100 points distribution	90%	Common sense suggests if 90% or more participants agreed on a core set, one could claim a consensus, albeit without unanimity.
Van Der Heijde (1997)	NGT	100 points distribution	Outcomes chosen by at least 3 out of 4 groups	Not reported
Smolen (1999)	NGT	100 points distribution	85%	Not reported
Bowman (2001); Pillemer (2005)	Workshop	Not reported	90%	Not reported
Ruperto (2003)	Delphi and NGT	Rank their top 10 choices (in order of importance) – Delphi.	Minimum of 80% (NGT at meeting)	Not reported
Cross (2005)	Delphi	Ranked as: Should be omitted, Not useful, Neutral, Useful and Essential	70%	Not reported
Mease (2005) Mease (2007)	Workshop	Asked to rank which domains they thought were most important, and second and third most important to measure.	50% 70%	Not reported
Devane (2007)	Delphi	5 point Likert scale: 1 = of no importance; 2 = of some importance; 3 = of moderate importance; 4 = very important, and 5 = extremely important	70%	Not reported
Gladman (2007)	Meeting	Not reported	72.30%	Not reported
Dent (2008)	Delphi and meeting	6 point scale: 1= agree strongly (A+); 2=agree moderately (A); 3=just agree (A)); 4=just disagree (D)); 5=disagree moderately (D); 6= disagree strongly (D+)	75%; *a priori*	Not reported
Khanna (2008)	Delphi	9 point scale: Median score 1-3 inappropriate (complete lack of consensus), 4–6 uncertain (some consensus), 7–9 appropriate (good to excellent consensus).	66%; *a priori*	We chose 66% instead of the traditional 80% as there were only nine members in the Steering Committee and an 80% consensus would have required agreement among eight of nine

				experts. This latter was considered an unrealistic and overly stringent requirement.
Taylor (2008) Schumacher (2009)	Delphi Workshop	Seven point scale: 1=definitely necessary to 7=definitely not necessary	UCLA/RAND disagreement index, whereby values of less than 1 indicated agreement. This index is essentially calculated from the 30th and 70th percentile of the respondents' ratings, adjusted for symmetry between the central point of the interpercentile range and the mid-point of the rating scale. 70%	Not reported
Douglas (2009)	Delphi and NGT	Score from 1 (least appropriate) to 9 (most appropriate Delphi. Outcomes with a score of 6 or greater were considered in the committee meeting (NGT).	80% (NGT)	Not reported
Schmitt (2011) Schmitt (2012)	Delphi Meeting	9 point scale: 1–3 = not important; 4–6 = equivocal; and 7–9 = important.	60% of all members of at least 3 stakeholder groups; *a priori* Less than 30% of the voters disagree; *a priori*	Not reported
Bennett (2012)	Delphi	Rank top three outcomes	Two or more of nine stakeholders ranked it among top 3 outcomes	Not reported

Heiligenhaus (2012)	Delphi and meeting	5 point scale, 1 highest importance and 5 lowest importance	100%	Not reported
Salaffi (2012)	Delphi	Likert 3 point scale (patients): 1=not relevant, not important; 2=not very relevant, not very important; 3=very relevant, very important Likert Four point scale (clinicians): 1=not relevant, unimportant; 2=not very relevant, not very important; 3=very relevant, very important; 4=highly relevant, extremely important	70%	Not reported
Lynch (2013)	Survey	5 options: Primary importance, Secondary importance, Not important/do not use, Indifferent, Unfamiliar with measure. 'Primary Importance' and 'Secondary Importance' were operationally defined as positive.' Not Important/Do Not Use' was operationally defined as negative. 'Indifferent' was operationally defined as having no impact on the overall value of a measure.	80%	An 80% threshold has been used previously to identify consensus in surveys of orthopaedic surgeons with a different rating scale.
Smaïl-Faugeron (2013)	Delphi	5 point scale: 1, no importance; 2, some importance; 3, moderate importance; 4, very important; and 5, extremely important	70%	Not reported

3.4.9 People involved in selecting outcomes

Table 9 shows the participant groups that were included in these studies. The types of people who are regarded as (or determined to be) key to developing a COS will likely vary between clinical areas, but two stakeholder groups that are likely to be important to all COS are clinical experts and patients. Where the types of people involved were described in the studies in this review, almost all COS included clinical experts (173/174 studies), but only 18% (31/174) included patients in the process.

Table 9: Participant groups involved in selecting outcomes

Participants category	Sub-category	N
Clinical experts (n=171)	Clinical experts*	88
	Clinical research expertise**	67
	Clinical trialists/Members of a clinical trial network	10
	Others with assumptions***	52
Patient representatives (n=31)	Patients	20
	Carers	7
	Patient support group representatives	9
	Service Users	2
Non-clinical research experts (n=54)	Researchers	26
	Statisticians	20
	Epidemiologists	11
	Academic research representatives	4
	Methodologists	6
	Economists	3
Authorities (n=40)	Regulatory agency representatives	31
	Governmental agencies	12
	Policy makers	4
	Charities	1
Industry Representatives (n=32)	Pharmaceutical industry representatives	29
	Device manufacturers	2
	Biotechnology company representatives	1
Others (n=70)	Ethicists	1
	Journal editors	2
	Others**** (with known participants)	15
	Others with assumptions***	52
No details given (n=26)		26

** Clinical experts includes multiple descriptions*

***16 studies, participants described as 'researchers/investigators' or 'academic researchers'*

**** 52 studies with clinical input but unclear about involvement of others*

***** Workshop/meeting participants (*5), subcommittee/committee (*2), guidelines panel, military personnel, moderator and audience, representatives from EORTC, members with expertise in information technologies, informatics, clinical registries, data-standards development, expertise in vaccine safety, malaria control and representatives from funding agencies/registration authorities, and donor organisation, members of the Rheumatology Section of the American Academy of Pediatrics, the Pediatric Section of the ACR, and the Arthritis Foundation, the diagnostic radiology and basic science communities, and from individuals conversant with functional and quality of life (QOL) assessments, comparative effectiveness research, and cost/ benefit analysis*

Patient representatives were identified most commonly via medical institutions (n=10), and four of these studies also used a charity or support group to identify patient participants. However, the majority of studies that included patient representatives did not describe how they were identified (18/31 studies, 58%). The number of patients that they included was not reported in 11 studies. A description of the methods used, the number of patients included and the proportion of the total participants this represents is given in Table 10 (n=20). It was not always clear what part of the COS development process patients were included in (12/31 studies, 39%). In 12 studies, patients were included in generating a list of outcomes and prioritisation of outcomes, and the remaining seven studies included patients in the prioritisation of outcomes stage only. Only three studies provided some description of how the material for explaining outcomes was developed for this group of stakeholders. In two studies, clinicians explained verbally what was meant. One of these studies, and an additional study, also carried out a pilot phase where patient representatives were asked whether the questions or items were easy to understand and appropriate, and the wording was then refined accordingly.

Table 10: Patient participation detail

	Method	**Total number of participants n**	**Number of patient representatives n**	**% patient**
1	Delphi (mixed panel) - *Number of rounds not clear, all took part in all rounds*	10	1	10%
2	Consensus Process (guidelines for trials) - review of RCTs and open discussion	6	2	33%
	Survey (mixed)	461	Not reported	Unknown - *Of 335 suggestions, 68% were from patients*
3*	Workshops (mixed)	OMERACT 6: 57 OMERACT 7: 179	OMERACT 6: 11 OMERACT 7: 19	OMERACT 6: 19% OMERACT 7: 11%
	Meeting (mixed)	OMERACT 8: 80	OMERACT 8: 20	OMERACT 8: 25%
4**	Interviews (patient only)	23	23	100%
	Nominal Group Technique (patient only)	26	26	100%
	Postal survey (patient only)	254	254	100%
5	Focus groups (mixed)	27	12	45%
	Rating exercise (mixed)	38	19	50%
6	Surveys (parents and children) and Delphi (clinicians) - *same study*	Round 1: 95 Round 2: 93	Round 1: 49 Round 2: 50	Round 1: 52% Round 2: 54%
7	SR and survey (mixed)	12	6	50%
	Delphi (mixed)	46	6	13% (same for all 3 rounds)
	Meeting(mixed)	43	5	12%

8	Delphi (mixed)	Round 1: 83 Round 2: 75 Round 3: 68	Round 1: 44 Round 2: 38 Round 3: 32	Round 1: 53% Round 2: 51% Round 3: 47%
9***	Focus groups (patient only)	31	31	100%
	Survey(patient only)	959	959	100%
10	Focus groups (patient only)	48	48	100%
	Delphi (patient only) *Did separate patient and researcher Delphi*	Pretest: 100 Round 1: 73 Round 2: 84	Pretest: 100 Round 1: 73 Round 2: 84	100%
	OMERACT 9 module (mixed)	not clear	not clear	Unknown
11	Rating exercise (mixed)	13	3	23%
12	Delphi (mixed)	Round 1: 218 Round 2: 173 Round 3: 152	Round 1: 9 Round 2: Not reported Round 3: 5	Round 1: 4% Round 2: Unknown Round 3: 3%
13	Advisory panel meeting (mixed)	11	2	18%
14	Step 4 - survey and meeting (mixed) Step 6 - Delphi (mixed - round 3 only related to outcomes - previous rounds related to priority research questions)	Step 4 - 6 Step 6 - 9	2	Step 4 (33%) step 6 (22%)
15	Delphi (mixed) - *rounds not reported*	338	86	25%
16	Consensus conference (mixed)	36	2	6%
17	Survey (mixed)	136	5	4%
18	Workshop (mixed)	39	2	5%
19	Workshop (mixed)	23	1	4%
20	Workshop (mixed)	23	2	9%

**COS had already been developed without patient input, so this work done to elicit patient opinion*
******* Patient core set*
****COS had already been developed without patient input, so this work done to elicit patient opinion*

Table 11 shows the participants' geographical location according to continent, as reported in the articles, as well as the median and range of number of countries included. In 34 studies, locations for participants other than the lead contact/participating authors were not provided. The geographic locations of participants were predominantly North America (n=164; 83%) and Europe (n=150; 76%). The remaining continents were represented in less than a quarter of studies; Australasia (n=47; 24%), Asia (n=40; 20%), South America (n=23; 12%) and Africa (n=13; 7%). The number of countries involved in the development of a COS ranged from 1 to 46 (a median of 4).

Table 11: Participants' geographical location

Continents	N (%)	Median and range of number of countries
North America and Europe [1] *	56 (28)	4, 2-25
North America [2]	44 (22)	1, 1-2
Europe [3]	32 (16)	2, 1-14
North America, Europe and Australasia [4] *	13 (7)	7, 3-25
North America, Europe and Asia [5]	11 (6)	9, 5-14
North America, Europe, Australasia, Asia [4]	10 (5)	11, 6-15
North America, South America, Europe, Australasia and Asia [4] *	10 (5)	16, 5-21
North America, South America, Europe, Australasia, Asia and Africa *	4 (2)	26, 8-46
North America, Europe, Australasia and Africa [5] **	3 (2)	8, 3-17
North America and Australasia	2 (1)	3, 3
North America, South America and Europe	2 (1)	10, 9-11
North America, South America, Europe and Asia	2 (1)	11, 7-15
Australasia	1 (<1)	2
North America, Europe and Africa	1 (<1)	10
North America, South America, Asia and Africa	1 (<1)	5
North America, South America, Europe and Australasia	1 (<1)	11
North America, South America, Europe and Africa	1 (<1)	7
North America, Europe, Australasia, Asia and Africa	1 (<1)	15
North America, South America, Europe, Australasia and Africa	1 (<1)	8
North America, South America, Europe, Asia and Africa	1 (<1)	18
Europe and Australasia	1 (<1)	2

Besides the lead contact or participating authors, other participants' locations were not stated/known ([1] – 15 studies, [2] - 9 studies, [3] - 7 studies, [4] - 2 studies, [5] - 1 study)

** In 6 studies, OMERACT participants' information was extracted from the introductory paper*

*** In 1 study, participants' location was based on where they had graduated*

3.4.10 Implementation and future updates

Two papers explicitly suggested or discussed plans for the implementation of their recommendations. Giacoia [80], which is linked to three COS studies, stated that their proposed recommendations will be considered as a requirement by a governmental agency (NICH):

> "The study-design frameworks and/or end points proposed by the initial groups will be considered by the NICHD (National Institute of Child Health and Human

Development) in the requirements for BPCA (Best Pharmaceuticals for Children Act) contracts for the study of off-patent drugs in the newborn population."

Cranney [81] stated that the proposed core measures were being submitted to key international research groups in osteoporosis to promote their acceptance and implementation:

> "These suggested core measures are being submitted to the key international groups involved in osteoporosis research to comment and encourage the widest possible acceptance and implementation of these tools in future randomized clinical trials."

167/198 COS (84%) did not report plans to update their recommendations. Thirty-one (16%) included some information about updating their COS recommendations. In six studies, the reported planned update was in relation to how to measure the outcomes identified in the existing study, rather than reviewing the outcomes to measure. Two out of the 31 studies reported that it was hoped that an interested group of participants would keep the development of the core set ongoing, but did not provide any further detail. A further two studies hinted that there would be an on-going review of their recommendations, but did not say anything more specific; one described a "continued re-evaluation" and the other "ongoing development" of the core set. 5/31 of the studies stated that their recommendations would be discussed and evaluated at future OMERACT conferences. Three studies reported that an International group had been founded to maintain and review the core set, but none of these stated specific plans or time frames for this. One study reported that the recommendations would be re-visited, as appropriate, as new evidence became available. Twelve studies reported a specific time frame to update their recommendations, which ranged between 1 year and 5 (Table 12).

Table 12: Range of years stated to review COS

Years to planned update	Number of studies
1	1
2	4
3-4	2
4	1
4-5	1
5	2
No more than 5	1

As per the inclusion criteria of the review, all of the included COS (N=198) made recommendations about the domains and/or outcomes, in other words, what to measure. Of the total, 75 (38%) made additional recommendations about how to measure those outcomes, that is on how to measure the suggested core outcomes. 5/75 studies recommended outcomes and instruments in a two stage process: 1) determination of the domains/outcomes; and then 2) determination of the instruments/ measurements. The majority of studies (70/75 studies; 93%) determined the outcomes and instruments for measuring those outcomes together and it was difficult to separate the two processes. 63/75 recommended specific instruments, such as SF-36, spirometry or HAQ-DI. The remaining 12 did not specify the instruments, rather provided general recommendations, such as use of a headache diary, health assessment questionnaire or visual analogue scale for pain. Although no recommendations were made for how to measure outcomes in the remaining 123 studies (62%), 54 of these (44%) did include discussion about this.

3.5 Discussion

This study provided the first complete assessment of COS that had been developed to standardise the outcomes being measured and reported in health research. The review identified 198 studies, in a range of health areas, and demonstrated that there has been a rapid increase in the development of COS over recent years. The studies identified in this review have been included in the COMET database.

Although a wide range of health areas were identified in this review, it is clear that some are more active in this field than others. This review allows the identification of areas where COS may be lacking, and these gaps provide future opportunities for COS developers. Developers need to define the scope of the COS at the outset in terms of health condition, population and types of interventions [38]. This review suggests that this has not always been done or is not described adequately in the reports, which also suggests a need for better reporting of studies of COS development. Furthermore, there may be other important areas of scope to consider that were not included in this review, such as the staging of the disease.

A striking aspect of the results is the infrequency with which patients have been included in the development of COS. Clinical trials are undertaken to establish whether interventions work and are safe for patients, so it is critical to include outcomes that they consider to be

important. We found that only 16% of studies (31/198 studies) included patient representatives in the development process, highlighting a need to find ways of engaging this group of stakeholders in particular in future projects, as well as other stakeholder groups who would be relevant to the COS. Most of the included studies included participants from more than one continent, but were dominated by North America and Europe. COS developers should consider including collaborators from other places as well; especially if a COS is to be applicable to, and adopted across, international settings.

3.5.1 Strengths and limitations of the review

I developed the search strategy in an iterative and methodological way to be highly sensitive, so that as many potentially relevant studies as possible were retrieved. Although every attempt was made to capture all relevant studies, a consequence of the lack of consistent indexing could be that some relevant studies were missed, along with studies that have been reported in journals and other places that were not indexed in the databases searched. We carried out hand-searching activities to try and minimise this. The search was performed in multiple databases, but these do have a bias towards research from North America and Europe. However, future efforts to identify COS and to minimise potential waste through unnecessary duplication would be for the bibliographic databases to introduce an indexing term to make them easier to find. Another limitation is that we were unable to undertake a formal quality assessment of the included studies. This is because defining the quality of a COS is not straight forward, and no validated way of doing this has been developed to date. There is an urgent need to develop such an instrument, not least to help users appraise the quality and relevance of a COS to their research and practice. This is discussed further in Chapter 8.

Finally, it is worth noting again that the first step in COS development is typically 'what to measure', which is the focus of this review; while the 'how' and 'when' usually come later. In this review we only included studies that addressed the first part of the process but, as an aside, of the 198 studies included in this review, 75 (38%) contained recommendations about how to measure the outcomes in the COS. Although it is not the focus of this study, it is vital to note that this is an important stage of COS development. Work is currently ongoing to identify how to select outcome measurement instruments for outcomes included in a core set [52].

The outcomes that were included in the 198 COS included in the review were extracted. I intended on classifying these outcomes in a way that could be more meaningful to allow us to make comparisons and look for differences. However, when we attempted this exercise I naturally started by looking to see what current ways of classifying outcomes exist. As highlighted in Chapter 1, there are various ways to classify outcomes. This, combined with the high number of studies included in the review, meant it was too complex to categorise outcomes as it was unclear which of these various frameworks was most suitable. The issue of categorising outcomes is elaborated on further in Chapter 8.

3.5.2 Implications

This systematic review provided the foundations for an online resource (www.comet-initiative.org). This is a freely accessible, publically available, searchable database that shows what work has been done in a particular health area. It will help to avoid unnecessary duplication of efforts and reduce waste in the production and reporting of research. Studies identified through this extensive review, which were not already included in the COMET database, have been added and an annual search of the literature will take place to keep the database current. The ready availability of COS should make it easier for researchers to design new trials. For example, the SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) guidance for protocols of clinical trials [56], includes a statement encouraging trial investigators to ascertain whether a COS exists relevant to their trial, and if so, to include those outcomes in their trial. The findings from this systematic review will help trialists to do this. Furthermore, applicants to the NIHR HTA programme in the UK, the Health Research Board in Ireland and the charity Arthritis Research UK, are now encouraged to consider COS when seeking funding for new trials. The COMET database will provide a resource for this.

The implications of this review go beyond clinical trials; with developers of 11% of the COS identified noting that they intended their recommendations for clinical practice, as well as health research. Furthermore, the National Institute for Health and Care Excellence (NICE) in the UK develops guidelines to help health and social care professionals deliver the best possible care based on the available evidence and, since 2009, has used standard criteria (Grading of Recommendations Assessment, Development and Evaluation, GRADE) to assess the quality of the evidence by outcome, rather than by study. In addition to these methods, NICE now emphasises checking of the COMET database in their guideline development

process. This highlights the importance of the results of this review for the improved delivery of healthcare.

3.5.3 Future work

The credibility of a COS depends on both the use of sound methodology in its development and transparent reporting of these methods. This review has highlighted the need to improve the standards of reporting, and the COMET Initiative is now undertaking this task. Work will build on the preliminary checklist [38] based mainly on discussions among the COMET Management Group. It will follow the strategy proposed in EQUATOR guidelines [82] involving five major phases: initial steps, pre-meeting activities, face-to-face consensus meeting, post-meeting activities and post-publication activities. This is discussed in more detail in Chapter 8.

This review highlighted that COS have been developed in a range of ways and studies are of variable quality. There is currently no accepted gold standard method, and no way of determining or assessing the quality of COS studies; this review has highlighted the need for such an instrument. A quality assessment instrument for studies developing COS will need to use criteria that are valid and reliable so that COS developers and users can assess the quality of a COS, helping in the decision about whether a COS is good enough to be adopted on, in some cases, in choosing between COS. This is discussed further in Chapter 8.

Categorising outcomes was a major challenge of this review. An outcome classification system that would allow this would also be helpful to future COS developers to help them think about potentially relevant outcomes. As such there is the need for a more formal review of the existing frameworks and schemes for doing this. An appraisal of existing methods would allow us to see which, if any, is most suitable for categorising outcomes. This is discussed further in Chapter 8.

3.6 Summary

In conclusion, this review identified 198 studies that have addressed the development of COS for measurement and reporting in clinical trials. I have highlighted future areas of research, including the need for methodological guidance for COS development, and methods for engaging key stakeholder groups, in particular patient representatives. Finally,

the review has shown that it is not always possible to identify key features of the development of a COS from the published report, highlighting a need for better reporting and indexing of COS development studies. The work done for this review has brought together the existing research in a single place, and provides a basis for improving standards for ongoing and future work to develop COS.

The rationale for methodological decision making, including choice of methods, choice of stakeholder groups and definitions for consensus, has not been well documented to date. By exploring this we will better understand why COS developers have done things so far, enabling us to better inform how COS developers might do this work in the future. Furthermore, it would seem that COS developers have not reported their plans regarding the implementation of their recommendations, or plans to review or update them. These may not have been a priority for COS developers, or may have been beyond the scope of these papers. Further work is required to explore these important issues. Chapters 5 to 7 describe work undertaken to understand COS developer's experiences of COS. Firstly, Chapter 4 will describe additional data, from the systematic review, about the most commonly used methods by ongoing studies. It is important to know more about the methods that are being used presently in order to inform current practice.

Chapter 4: Methods used to develop core outcome sets

4.1 Background

The systematic review described in Chapter 2 and 3 identified 198 studies, in a range of health areas, and demonstrated a rapid increase in the development of COS over recent years. The systematic review highlighted that a range of methods had been used, in a variety of ways, to develop COS in these published studies. The first part of this chapter will provide a description of general COS methods identified in the systematic review.

While the systematic review provided information about the range of methods that had been used to develop COS historically, we need to know more about the methods that are being used presently in order to inform current practice. At the end of 2014, the COMET database included a total of 57 studies listed as ongoing COS studies. The most common combination of methods used by these ongoing studies was a systematic review, together with a Delphi survey, and sometimes followed by a consensus meeting. This chapter will therefore describe additional data, from the systematic review, about studies that used these specific methods, so that we can learn more about how these methods have been used previously.

4.1.1 Overview of methods used in core outcome set development

The methods described in this section were used in studies identified in the systematic review of COS. A summary of the characteristics of each method is provided in Table 13.

Table 13: Characteristics of methods used in the development of core outcome sets

	Group interaction	Elicit individual responses from all participants	Face to face	Non face to face (e.g. telephone, internet, email)	Participant anonymity
Group discussion	✓	X	✓	X	x
Focus group	✓	✓	✓	X	x
Consensus development conference	✓	X	✓	X	x
Delphi Technique	x	✓	✓	✓	✓
Questionnaire survey	x	✓	x	✓	✓
Nominal Group Technique (NGT)	✓	✓	✓	X	x
Individual interview	x	✓	✓	✓	✓
Systematic review	x	X	x	X	x

Group discussion

Unstructured group meetings, such as committees, are largely open, unstructured and informal with few rules, procedures or guidelines. Interaction between participants is not structured. These group discussions, which are often termed "free discussion" or simply "consensus" groups, bring together a group of people to discuss a problem with the ultimate aim of reaching consensus.

The dynamics involved in groups may adversely impact on the decision making process and quality of decisions reached. These include that the desire to reach agreement may override the true result; and dominant individuals or social factors may affect group performance [83]. Furthermore, although this method would potentially allow the presence of multiple or different stakeholders, the lack of formal process and structure may not ensure that each stakeholder's view is elicited.

In semi-structured or structured group discussion, the discussion is often steered by a facilitator. This more formal structured process may eliminate some of the negative aspects of unstructured group decision making [83]. Semi-structured discussion techniques might include discussion at conference meetings and workshops. Focus groups are one example of structured group discussion.

Focus group

Focus groups are a form of group interview, in which group members meet face to face together in one place, to discuss a particular issue under the direction of a facilitator. Groups are usually small, typically six to twelve people, and usually last between one and two hours. Each participant may also complete individual questionnaires to enable the researcher to compare what is said in public and private.

Focus groups have the advantage of making use of group dynamics to stimulate discussion, gain insight and generate ideas to pursue a topic in greater depth [84]. It has been suggested that focus groups allow the interviewer to encourage participants to explore the issues of importance to them, and pursue their own priorities [85]. There are also the associated downsides of such group dynamics including the potential for vocal individuals to dominate discussion, which may limit the usefulness of the data. While this method allows for multiple/different stakeholders to be involved simultaneously, small group numbers may make it challenging to make sure that all key stakeholder groups are represented and representative. It may therefore be necessary to conduct multiple focus groups, which is likely to be time consuming and expensive.

Consensus development conference

In the 1970s,The National Institutes of Health (NIH) set up the NIH Consensus Development Programme in the US with the aim of bringing together people to seek general agreement on the safety and efficacy of use of various medical procedures, drugs and devices [86]. As part of this, they held consensus development conferences that they argued were preferable to traditional conferences. Traditional conferences address the status of research on a particular topic with limited discussion amongst individuals with diverse viewpoints, while consensus development conferences focus on the application of research and the objective is to provide resolution, including in situations where no clear answers are possible in the existing environment. One of the main reasons for initiating consensus development conferences in the UK was to promote change in health care policy and practice [87].

Key features of consensus development conferences include:

(1) *Background reports are prepared* – individual experts or task forces are commissioned to compile summaries of current knowledge

(2) *Audience participation is encouraged* –all NIH consensus development conferences were free and open to the public

(3) *Panel to assess the evidence presented* – panel members should be carefully selected to represent the range of individuals and organisations with expertise and interest.

The conference begins with a session for individuals or representatives of task forces to present information. Comments by panellists may follow, and there is then time for questions and comments from the audience. The panel then convenes in an attempt to reach agreement. In the final session, the consensus statement(s) are presented to the audience, and there is again time for comments, and endorsement. The consensus statement may then be revised after the conference for publication. This may also be circulated to a "larger" group after the conference to seek a broader consensus.

Consensus development conferences have been criticised for providing little information on how consensus is defined, having a lack of formal criteria for making decisions and an absence of formal voting, as well as variations in the process. These variations include the nature of the audience and who is deemed the primary target of the consensus outputs; the pre-panel process including selecting the chair, panel members and preparation of background information; the panel meeting stage including meeting activities and the group process by which consensus is actually achieved [88]. In a critique of the consensus development conference following a conference about early melanoma, Ackerman reports that one subject discussed was not alluded to in the statement issued by the panel. Akerman goes on to suggest that the volume of information together with unreasonable time constraints may be problematic for the panel, no matter how competent or representative the composition [89]. Critics have also suggested that no new data are generated by consensus conferences and that innovation in research may be stifled or ideas restricted [87].

Delphi technique

The Delphi technique was designed as a method of improving decision-making by eliciting, formulating and refining group judgements. It is a method of reaching consensus, employing the rationale that "n heads are better than one," when dealing with conflicting scientific evidence or uncertainty [90]. The Delphi technique was originally developed as a forecasting tool for technological developments, but its early application extended to policy

decisions in education, public transportation and public health [90]. Its application today remains diverse, but it is being increasingly used in research studies aiming to achieve consensus about the outcomes that should be measured in trials of specific conditions.

Dalkey described three key features of the Delphi technique:

(1) *Anonymous response* – opinions of the group are obtained by formal questionnaire
(2) *Iteration and controlled feedback* – interaction is effected by a systematic exercise conducted in several iterations, with carefully controlled feedback between rounds
(3) *Statistical group response* – the group opinion is defined as an appropriate aggregate of individual opinions on the final round.

The advantages of this technique are that it helps to minimise the likelihood that the group is influenced by the views of dominant or senior individuals, to avoid irrelevant communication, and to minimise or avoid the influence of group pressure to conform. It is also generally fast, inexpensive and easy to understand [91]. Improvements in global communication (e.g. internet) have made it feasible to use the Delphi technique to involve geographically distant participants, and in larger numbers, and from different/multiple stakeholder groups, than may be possible in face to face situations [59].

The Delphi technique has been subject to criticism and the validity of the technique has been questioned. One critique of the Delphi technique questions whether a group dilutes the opinion of "the real expert" on the particular question or uncertainty that is being considered [92]. The technique has also been criticised for being open to interpretation. For example, the use of the word "appropriate" in the description by Dalkey raises methodological questions about how "appropriate" is decided. This inevitably leads to variation in how the Delphi technique is conducted, and could therefore also influence the quality of the decisions reached. Furthermore, the question of how an "expert" is defined has also been criticised as this is an arbitrary term [91, 93]. The title of 'expert' is misleading, and refers to individuals who have knowledge of, or experience with, a particular topic, and the question of definition relates to the selection procedures used for selecting panellists. It has also been suggested that anonymity may lead to lack of accountability of the views expressed, resulting in hasty or ill-considered judgements [93], and that the feedback process may in fact lead participants towards superficial or premature conformity rather than consensus or true agreement [91]. Furthermore, the

existence of consensus does not mean that the "correct" answer has been found, as "misinformation may be aggregated into a less reliable opinion"[90]. As one of the most highly utilised methods, it may be that the Delphi technique has been subject to more critical scrutiny than the other methods described here.

In a review of 150 studies using the technique for scientific and technological forecasting in a range of areas, such as education, communication technology and medicine, Sackman found that many variants existed with some departing widely from the original Delphi procedure, and he concluded that there was an absence of an agreed, universally accepted, working definition of Delphi. Despite these criticisms, the technique has continued to be widely used. More recently, however, the technique has also been criticised for there being no consistent method for reporting findings of such studies, and it is suggested that such standards might help readers to judge the reliability of the method and results obtained [94]. A systematic review of studies that used the Delphi technique to determine which outcomes to measure in clinical trials, similarly concluded that there was variability in both methodology and reporting of studies using Delphi for this purpose [59]. This resulted in recommendations to improve the quality of studies that use the Delphi process for determining outcomes to use in clinical trials, including the recommendation that patients and clinicians be involved, researchers and facilitators avoid imposing their views on participants, and attrition of participants be minimised.

Questionnaire survey

Questionnaires are often used as a means of collecting information about people's knowledge, opinions and attitudes. Questionnaires can be used as the sole research instrument, but are often used within a mixed methodology study. Questionnaires may be validated or not, may use open ended or closed questions and can be presented in various ways [95]. They are used in face to face, postal and telephone surveys. The main strengths of a structured questionnaire include the ability to collect unambiguous and easy to count answers. The method leads to economical data collection, so large samples can often be included [84].

Their weakness is that limited response choices may not be sufficiently comprehensive, and not all answers may be easily accommodated, so responders may be forced to choose an answer that might not fully represent their view. Furthermore, nuances and contextual factors are difficult to capture. Although the advantages of a self-administered

questionnaire or postal questionnaire may be a reduction in social desirability and interviewer bias, this technique is not recommended for complex issues or long questionnaires as the interviewer is unable to clarify questions or to probe [84]. It is also noted that it might be possible for respondents to read all the questions before answering any, with the possibility of answering in any order they wish, which may affect their response.

Nominal Group Technique (NGT)

The NGT was developed in the context of committee decision making, for the purpose of structuring interaction within a group [96]. The purpose of NGT is to generate information in response to an issue that can then be prioritised through group discussion [97], and since its development, NGT has continued to be widely used in a range of fields including education and health. There have been an increasing number of health-care studies that use NGT, for example to develop a list of problems that people with type 2 diabetes associate with sleeplessness [98].

Firstly, each participant records their opinion independently. Then, in a "round-robin" format by means of a facilitator, the ideas are listed from each individual in turn and listed in front of the whole group. Each idea is then discussed, in turn, by the group. Each member of the group is then asked to vote or rate each idea, before coming together to discuss. Further discussion and voting may take place, and the final group opinion is compiled. The key feature of NGT is structured face to face meetings. It allows disparate ideas to be expressed and collated, with a view to identifying areas of consensus and establishing priorities for change, with the collaborative nature of NGT said to increase the participants' ownership and accountability [99]. It enables participants to become fully involved in the decision making, and prevents groups from focussing on one particular area at the expense of others.

One advantage of the NGT is that the process provides both quantitative data in the voted upon priorities to quantify variation, and qualitative data in terms of descriptive discussion to describe variation. It also ensures that each participant's view is represented, allowing inclusion of multiple/different stakeholders; helping to minimise the likelihood of the group being influenced by the views of dominant or senior individuals. Practical advantages of NGT include time and resource efficiency, as participants' input is usually limited to a single meeting [97]. However, this may also be seen as a disadvantage as all participants must be

available to convene at the same time and place. The often immediate dissemination of results to the group involved, may also promote satisfaction with participation [97].

Individual interview

The interview is the most widely used method of producing qualitative health research [100]. It is directed towards the researcher's particular needs for data. Interviews are often classified according to how far the researcher directs the interview. In structured interviews, the researcher directs the interview completely, and follows a specified set of questions in a particular order. These are typically used in survey designs. At the other end of the scale are informal interviews. These are often more conversational in nature and are likely to occur fortuitously, leading to a more opportunistic gathering of data. Between these extremes is the semi-structured interview, also referred to as narrative and in-depth. In this type of interview the researcher will set the topics covered and ask specific yet open ended questions, but the interviewee's response directs the interview and the kinds of information produced [100].

Advantages of individual interviews include providing the opportunity to explore topics in depth, allowing the interviewer to explain or clarify questions thereby increasing the likelihood of useful responses, and allowing the interviewer flexibility in administering the interview to the particular individual or circumstance. One main advantage of this method is that it allows each participant the opportunity to express his/her own viewpoint. However, there are also disadvantages to this method. Individual interviews can be both expensive and time consuming [84], and there are other practical challenges including social, environmental and cultural factors [100]; and there is also the possibility that both interviewer and interviewee can influence the process. Flexibility across interviews could also result in inconsistencies across interviews.

Systematic review

A systematic review is a way to bring together the evidence to answer a specific research question [12]. It will usually fit pre-specified criteria and is done by combining the results of several studies. The procedure is designed in a way to minimise bias and uses transparent processes to find, evaluate and synthesise the relevant research. The Cochrane Collaboration (www.cochrane.org) strives to produce high-quality relevant systematic reviews and is the benchmark for high-quality information about health care. The Cochrane Handbook Systematic Reviews of Interventions states the following five criteria that define

a systematic review. Although a systematic review of outcomes differs to a review of interventions, the same underlying systematic review principles apply:

- a clearly stated set of objectives with pre-defined eligibility criteria for studies;
- an explicit, reproducible methodology;
- a systematic search that attempts to identify all studies that would meet the eligibility criteria;
- an assessment of the validity of the findings of the included studies, for example through the assessment of risk of bias; and
- a systematic presentation, and synthesis, of the characteristics and findings of the included studies.

Systematic reviews differ from literature reviews in that literature reviews tend not to include a systematic search of the literature and may focus on only parts of the available data [101]. Although literature reviews can be informative and can provide an overview of the area, they may include an element of bias that systematic reviews should not.

Throughout Chapter 4, I will provided details of studies identified in the systematic review of COS (as described in Chapter 3), that used these methods in COS development.

4.2 Aims

To provide information from published COS about those methods that are used most commonly in present COS development; specifically systematic reviews, the Delphi Technique and consensus meetings (when done in combination with Delphi).

4.3 Methods

The full systematic review methods are described in Chapter 3. The following data was extracted specific to these methods:

Literature or systematic review:

1. Topic/focus of review
    a. Did the study include a review of instruments/definitions?

2. Type of studies included

3. Language of included studies

4. Date range searched

5. Databases searched

6. How they classified/grouped outcomes

7. The conclusions of the review

The Delphi technique:

1. Number of panels

2. Number of rounds

3. Administration of questionnaires

4. How was the original list of outcomes generated?

5. Information provided to participants before the first round

6. What was asked in each round?

7. Method to remove or reduce the outcomes

8. Composition of the groups

    a. Countries and continents represented

9. Level of anonymity

10. How consensus was reached

    a. Was Delphi the final method?
    b. Definition of consensus

The recommended checklist of study characteristics and results that should be reported in studies using the Delphi technique to determine outcomes [59] was used to assess the quality of reporting in these studies.

Consensus meeting:

1. Description of what happened during the meeting (discussion, presentation, voting)

2. Duration of meeting

3. Who conducted/led the meeting
    a. Who assisted

4. Participants
    a. Number and type of participants
    b. Countries and continents represented

5. Level of anonymity

6. How consensus was reached
    a. Definition of consensus

4.4 Results

4.4.1 Literature or systematic review

Sixty-five studies (33% of 198 included in the systematic review) carried out a literature or systematic review as part of their methods to develop a COS. This was done in combination

with another method in 54 of these 65 studies (83%). Table 14 summaries the type of review that was performed in each study.

Table 14: The type of review that was performed

Type of review	n
Outcomes only	7
Outcomes and outcome measurement instruments	23*
Outcomes and outcome definitions	7*[1]
Outcomes, outcome measurement instruments and outcome definitions	9
Outcome measurement instruments	1
Study design only	7
Study design including outcome measurement instruments	2*[1]
Study design including outcome definitions	3
Study design including outcome measurement instruments and outcome definitions	2*[1]
Methodological issues only	1
Methodological issues including outcome measurement instruments	1
Quality of reporting including outcome definitions	1
study design and quality of reporting including outcome measurement instruments	1

* *This was not explicitly stated in 4 of these studies, but assumed due to the discussion in the paper*

*[1] *This was not explicitly stated in 1 of these studies, but assumed due to the discussion in the paper*

The rest of this summary relates to the 46 studies that carried out a review of outcomes.

Eighteen studies (39%) did not state the type of studies they included in their review. Of those that reported the types of studies they included, clinical trials were the most frequent type of study to be included (24/28; 86%), one of which also included longitudinal, cohort studies, case control studies and consecutive case report series in addition to clinical trials. The descriptions of the included studies in the remaining four studies are provided in Table 15.

Table 15: Descriptions of other types of included studies (n=4)

Descriptions of included studies
Articles relating to the use or validity of asthma indicators
Existing data standards sources
Articles discussing research design, cohort, case control, and cross-sectional studies
Research cited in community acquired pneumonia guidelines

Twenty of the studies (44%) did not provide information about the language of included studies. Eighteen of the 26 that reported language restrictions (69%) included English studies only, and three included other language studies in addition to English (Italian; German and French; German, French, Spanish and Italian). Five studies did not apply any language restrictions to their search.

Seventeen studies (37%) did not state the date range searched. Six studies (21%) did not apply any date restrictions to their search. The number of years reported in the remaining twenty-three studies ranged between two and 59. Frequencies are provided in Table 16.

Table 16: Number of years searched (n=23)

Number of years searched	Frequency
Less than 5	1
5 to 9	3
10 to 14	5
15 to 19	2
20 to 24	5
25 to 29	0
30 to 34	0
35 to 39	4
More than 40	3

The number of databases searched was not reported for seventeen studies (37%). Two studies did not perform an electronic database search. Twenty-seven studies described which databases they searched (Table 17). The number of databases searched ranged between 1 and 8; frequencies are also provided in Table 17.

Table 17: Frequency of number of databases searched (n=27)

Number of databases	n	Databases searched
1	15	Medline (n=9)
		PubMed (n=4)
		Central Register of Controlled Trials (n=2)
2	6	Medline and Embase (n=2)
		PubMed and Central Register of Controlled Trials (n=1)
		Medline and CancerLit (n=1)
		Medline and Central Register of Controlled Trials (n=1)
		Medline and PubMed (n=1)
3	1	Medline, Cinahl and Embase (n=1)
4	0	
5	1	Medline, PreMedline, CancerLit, PubMed (National Library of Medicine) and Cochrane Library
6	0	
7	2	Cochrane Oral Health Group's Trials Register CENTRAL, Medline, Embase, Science Citation Index Expanded, Social Science Citation Index, Index to Scientific and Technical Proceedings, System for Information on Grey Literature in Europe (n=1)
		Cochrane Skin Group Specialised Register, the Cochrane Central Register of Controlled Trials in The Cochrane Library (Issue 4, 2009), Medline, Embase, AMED, PsycINFO, LILACS (n=1)
8	2	CINAHL (Cumulative Index to Nursing and Allied Health Literature), Embase, Medline, National Criminal Justice Reference Service (NCJRS), PsycINFO, Sociological Abstracts, The Cochrane Database, The Patient-reported Health Instruments (PHI) website (n=1)
		Medline, PubMed, Embase, PsycINFO, Cinahl, Web of Sciences, Cochrane Central Register of Controlled Trials, and Cochrane Database of Systematic Reviews (n=1)

Ten studies provided some detail about how they classified/grouped outcomes. The level of detail is variable, and is provided per study in Table 18.

Table 18: Methods for classifying/grouping outcomes (n=10)

Reference	Method for classifying/grouping outcomes
Duncan (2000)	Each outcome measure was classified into one of the following categories: death or, at the level of pathophysiological parameters (blood pressure, laboratory values, and recanalization), impairment, activity, or participation. Measures were classified according to the system used by Roberts and Counsell [102], which includes the Rankin/modified Rankin scale as a measure of activity rather than participation.
Sinha (2012)	Each outcome was grouped into one of the following six outcome domains, some of which were further divided into subdomains: 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. Reference given in support of this approach: Sinha et al 2008, 'A systematic review of studies that aim to determine which outcomes to measure in clinical trials in children' [58].
Broder (2000)	List developed by staff at institution (but no further detail).
Distler (2008)	The results of this literature search were discussed at the first meeting of the steering committee. Based on this discussion, a list of 17 domains and 86 tools was set up for the first stage of the Delphi exercise to define outcome measures for a clinical trial in PAH-SSc. Domains were defined as a grouping of highly related features that describe an organ, disease, function, or physiology (e.g., cardiac function, pulmonary function, and quality of life).
Devane (2007)	Outcome measures addressing similar dimensions or events were discussed by the team and collapsed where possible. For example, various modes of delivery/birth were presented as ''mode of birth (e.g., spontaneous vaginal, forceps, vaginal breech, caesarean section, vacuum extraction).'' This pilot tool was tested for clarity, with a sample of 12 participants, including 3 maternity care consumers, and subsequently refined.
Smaïl-Faugeron (2013)	"Because we expected a large diversity in reported outcomes, we grouped similar outcomes into overarching outcome categories by a small-group consensus process. The group of experts consisted of 6 doctors in dental surgery specialized in paediatric dentistry, including 3 clinical research investigators. First, the group identified outcomes that were identical despite different terms used across trials. Second, different but close outcomes (i.e., outcomes that could be compared across studies or combined in a meta-analysis) were grouped together into outcome domains. Finally, the group, with consensus, determined several outcome categories and produced a reduced-outcome inventory."
Merkies (2006)	In advance of the workshop, a list of outcome measures applied in treatment trials was prepared including their scientific soundness, WHO and quality of life classification (WHO classification reference is ICF).
Rahn (2011)	From this outcome inventory, the outcomes were organized and grouped into eight proposed overarching outcome domains. Categories were determined based on their applicability to all potential interventions for abnormal uterine bleeding and the physician expert group's consensus of their relevance for informing patient choices. Outcomes related to cost, resource use, or those determined by the review group to have limited relevance for assessing clinical effectiveness were excluded from categorization and further analyses.
Chow (2002)	Some detail but process not described - The endpoints employed in previous bone metastases trials of fractionation schedules were identified and listed in the first consensus survey under the following headings: 1. Pain assessments; 2. Analgesic assessments and primary endpoint; 3. Endpoint definitions; 4. Timing, frequency and duration of follow-up assessment; 5. When to determine a response; 6. Progression and duration of response; 7. Radiotherapy techniques; 8. Co-interventions following radiotherapy; 9. Re-irradiation. 10. Non-evaluable patients (lost follow-up) and statistics; 11. Other endpoints. 12. Other new issues and suggestions; 13. Patient selection issues.
Van Der Heijde (1997)	Grouped into patient assessed, physician assessed or physician ordered measures.

The results/conclusions of the reviews were not reported for 14 studies (30%). The most common conclusion was a lack of consistency in the measurement or reporting of outcomes (n=22). A summary of the conclusions drawn from the reviews is provided in Table 19.

Table 19: Conclusions from reviews, provided in 32 studies (not mutually exclusive)

Conclusion	N
No consistency in the selection/measurement/reporting of outcomes	22
Work is needed to develop a core set/agreement	9
Outcomes recommended	6
No consistency in how outcomes are measured (instrument, timing)	7
Outcome measures used in trials have limitations (e.g. not been validated)	7
No consistency in the definition of outcomes	4
Measurement (instruments) recommended	1

4.4.2 The Delphi Technique

Twenty-nine studies (15% of the 198 included in the systematic review) included the Delphi technique as part of their methods to develop a COS. Six studies used the Delphi technique only. The various combinations of Delphi used with other methods (n=23) are provided in Table 6 in chapter 3. Systematic reviews (n=10) or literature reviews (n=4) were conducted in 14 studies prior to the Delphi exercise. The Delphi technique was used in three studies to identify research questions or priorities related to trial design rather than prioritising outcomes, and are therefore excluded hereafter. A summary of the reporting quality of studies using the Delphi technique (n=26) is shown in Table 20.

Table 20: Reporting quality of the 26 studies using the Delphi Technique

		Before Sinha et al's guidelines (pre 2011)*		After Sinha et al's guidelines (2011 or later)		
Broad aspect of reporting	**Specific items assessed**	Clearly reported n (%)	Not clearly reported n (%)	Clearly reported n (%)	Not clearly reported n (%)	N/A n (%)
Size and composition of panel	Number of participants	19 (73)	0 (0)	7 (27)	0 (0)	0 (0)
	Types of participants (e.g. clinicians, patients)	19 (73)	0 (0)	6 (23)	1 (4)	0 (0)
	Proportion of each type of participant	18 (69)	1 (4)	6 (23)	1 (4)	0 (0)
	How participants were identified/sampled	19 (73)	0 (0)	7 (27)	0 (0)	0 (0)
Methodology of the Delphi	Administration of questionnaires (e.g. postal)	17 (65)	2 (8)	6 (23)	1 (4)	0 (0)
	How items were generated for first questionnaire	14 (54)	5 (19)	7 (27)	0 (0)	0 (0)
	What was asked in each round	15 (58)	4 (15)	6 (23)	1 (4)	0 (0)
	Information provided to participants before the first round	9 (35)	10 (39)	3 (12)	4 (15)	0 (0)
	How the overall group response was fed back	6 (23)	13 (50)	2 (8)	3 (12)	2 (8)[a]
	Level of anonymity	12 (46)	7 (27)	6 (23)	1 (4)	0 (0)
	A priori definition of ''consensus'' about whether an outcome should be measured	2 (8)	8 (31)	1 (4)	4 (15)	11 (42)[b]
	Were non-responders invited to subsequent rounds	9 (35)	6 (23)	3 (12)	2 (8)	4 (15)[c] 2 (8)[a]
Results	Number of respondents to each round	15 (58)	4 (15)	6 (23)	1 (4)	0 (0)
	Number who completed every round	13 (50)	6 (23)	4 (15)	3 (12)	0 (0)
	Results for each outcome in each round	0 (0)	19 (73)	5 (19)	2 (8)	0 (0)
	Group response for each outcome (final round)	11 (42)	8 (31)	6 (23)	1 (4)	0 (0)
	Distribution of response for each outcome in the final round	5 (19)	14 (54)	2 (8)	5 (19)	0 (0)
	List of all outcomes that participants agreed should be measured	10 (39)	0 (0)	5 (19)	0 (0)	11 (42)[b]

**Twelve studies were included in Sinha et al's review of Delphi studies. The systematic review in Chapter 3 identified an additional 7 studies in this time period.*

[a]Only one round of voting, therefore no feedback.

[b]Reaching final consensus was not the aim of the Delphi process.

[c]All participants responded to each round, so no discussion was made regarding non-responders.

Composition of the group

The number of participants varied from 8 to 338 with a median of 35. Fifteen studies included less than 50 participants (58%), which increased to 89% for less than 100. Only three studies included more than 100 participants. The size and composition of the groups is described in Table 21.

15 studies (58%) identified participants through one source only. Participants were most frequently identified through health professional (n=9) or clinical trial (n=8) networks. Five studies included participants known to the authors/facilitators; this was the sole method in one study. Other ways of identifying participants included published researchers (n=4), research groups (n=2) and recognised experts or professionals (n=2). Four studies identified participants (patients or clinical experts) via clinical settings (e.g. a clinic) and three studies used patient groups to identify participants.

Clinicians were involved in all Delphi Studies, although the percentage of clinicians was unclear in two studies. Twelve studies (46%) included 100% clinicians, and the remaining twelve studies included between 24 and 92%. Nine studies involved patient representatives in the Delphi study, which ranged from 10 to 76 %. The proportion of patients was unclear for one study. Eight studies included additional participants including methodologists, FDA/NIH representatives, industry representatives and journal editors. These are described in Table 21.

Ten studies conducted the Delphi in one Country only; eight in the US, one in the UK and one in Italy. The number of countries was unclear for two studies. The remaining 14 studies had between four and 46 countries involved, with a median of 11. A breakdown of these countries is provided in Table 21. Participants were involved from various continents. Most studies involved participants from North America (n=23; 92%) and Europe (n=16; 64%). Participants from Asia (n=10; 40%), and Australasia (n=9; 36%) were also represented. The least represented continents were South America (n=6; 24%) and Africa (n=4; 16%).

Table 21: Composition of Delphi groups

Study	Method of identification of the sample of participants	Number of participants	Number (%) of clinical professionals	Number (%) of patient representatives	Other types of participants included in the study (%)	Number of countries represented: Names (Continents)
White (1995)	Health professional network	12	11 (92)	0	Researcher n=1 (8)	1: USA (North America)
Broder (2000)	Speciality societies Staff at institution	10	9 (90)	1 (10)	0	1: USA (North America)
Basson (2000)	Specialty societies Key published researchers	19	19 (100)	0	0	5: Canada, Denmark, Italy, The Netherlands, USA (Europe and North America)
Miller (2001)	Clinical trial network Patient support group	70	Unclear	Unclear	0	14: USA, Hungary, Czech Republic, Sweden, UK, Israel, Mexico, France, Germany, The Netherlands, Canada, Guatemala, Italy, Korea (North America, Europe and Asia)
Ruperto (2003)	Clinical trial network	267	267 (100)	0	0	46: Argentina, Australia, Austria, Belgium, Brazil, Bulgaria, Canada, Chile, Costa Rica, Croatia, Cuba, Czech Republic, Denmark, Finland, France, Georgia, Germany, Greece, Hungary, Israel, Italy, Korea (South), Latvia, Lithuania, Luxembourg, Mexico, The Netherlands, New Zealand, Northern Ireland, Norway, Poland, Portugal, Russia, Saudi Arabia, Singapore, Slovakia, South Africa, Spain, Sweden, Switzerland, Tunisia, Turkey, UK, USA, Yugoslavia (North America, South America, Europe, Asia, Australasia, Africa)
Lux (2004)	Published researchers People known to facilitator	42	42 (100)	0	0	15: UK, Italy, USA, Argentina, Canada, Japan, Oman, Singapore, China, Philippines, Germany, Cuba, Switzerland, Malaysia, Thailand (North America, South America, Europe, Asia)
Cross (2005)	Recognized clinical experts	8	8 (100)	0	0	5: Ethiopia, India, Brazil, USA, Nepal (Asia, Africa, North America)
Lightfoot (2005)	Health professional network	35	35 (100)	0	0	1: USA (North America)
Lightfoot	Health	35	35 (100)	0	0	1: USA (North America)

(2005)	professional network					
Mease (2005, 2008)	Clinical trial network Local patients	96	23 (24)	73 (76)	0	1: USA (North America)
Taylor (2005)	Clinical trial network	32	32 (100)	0	0	10: countries not provided(North America, Europe, Australasia, Africa)
Devane (2007)	Health professional network Patient groups	218	147 (68)	24 (11)	Health service managers n=14 (6) Epidemiologists n=9 (4); Social scientists and lactation specialists (numbers of each group unknown) n= 24 (11)	28: United Kingdom, Canada, Australia, Ireland, United States, the Netherlands, Peru, Sweden, Germany, Northern Ireland, Singapore, Switzerland, Thailand, Zimbabwe, Saudi Arabia, Argentina, Georgia, New Zealand, Albania, Austria, India, France, Afghanistan, Brazil, Nigeria, Iran, and Denmark; one country missing (North America, South America, Europe, Asia, Australasia, Africa)
Dent (2008)	People known to facilitator Published researchers	13	12 (92)	0	Clinical trial methodologist n=1 (8)	9: Australia, USA, Canada, Denmark, UK, Japan, Norway, Italy, Belgium (North America, Europe, Asia, Australasia)
Distler (2008)	Clinical trial network	87	87 (100)	0	0	Unclear (North America, Europe, Australasia, Asia)
Khanna (2008)	Clinical trial network	50	50 (100)	0	0	15: USA, Italy, UK, Canada, Australia, Argentina, Hungary, France, Germany, Norway, Turkey, Ireland, The Netherlands, Sweden, Switzerland (North America, South America, Europe, Asia, Australia)
McGrath (2008)	People known to facilitator	26	17 (65)	0	FDA/NIH representatives n=5 (19) Industry representatives n=4 (16)	4: Canada, USA, UK, Sweden (North America, Europe)
Taylor (2008)	Clinical trial network	33	30 (91)	0	Industry representatives n=3 (9)	11: USA, Spain, Australia, Japan, Mexico, The Netherlands, Italy, UK, Russia, Thailand, China (North America, South America, Europe, Asia, Australasia)
Douglas (2009)	Health professional network	92	92 (100)	0	0	13 : USA, UK, Canada, Austria, Australia, Ireland, Demark, Spain, Turkey, South Korea, the Netherlands, Belgium and Iran (North America,

						Europe, Australasia, Asia)
Vargus-Adams (2009)	People known to facilitator Local clinics Health professional network	83	39 (47)	44 (53)	0	1: USA (North America)
Rahn (2011)	Research group	25	25 (100	0	0	1: USA (North America)
Schmitt (2011)	People known to facilitator Health professional network Patient self-help group	46	32 (70)	6 (13)	Regulatory agency representatives n=1 (2) Journal editors n=7 (15)	11: Australia, Brazil, Denmark, France, Germany, Italy, Netherlands, Sweden, Switzerland, United Kingdom, United States (Australia, South America, Europe, North America)
Bennett (2012)	Recognized professionals	9	4 (45)	2 (22)	Research funders n=2 (22) Epidemiologists/methodologists=1 (11)	1: USA (North America)
Heiligenhaus (2012)	Research working group	16	16 (100)	0	0	9: Germany, UK, Switzerland, Finland, Spain, Netherlands, Denmark, France, USA (Europe, North America)
Salaffi (2012)	Tertiary care unit	338	252 (75)	86 (25)	0	1: Italy (Europe)
Sinha (2012)	Health Professional Network Local clinics and hospital	95	46 (48)	49 (52)	0	1: UK (Europe)
Smaïl-Faugeron (2013)	Health professional network Published researchers Clinical trial network	62	Unclear	0	Unclear: International Authors of primary trials	France and international participants whose details are not provided (Europe, Other unknown)

Anonymity

It was unclear how three studies were conducted. Fourteen studies conducted the Delphi by email or internet (web-based), and three studies by post. Eight of these did not know the identity of other group members and were conducted anonymously. This was unclear in seven studies but it is assumed that they were conducted anonymously. There were three studies where the Delphi process was conducted solely at face to face meetings. In two of these meetings voting was anonymous, but this was unclear for one study. One study conducted the Delphi using email, and a meeting. Participants did not meet during

the Delphi, did not know the identity of the group members and answered anonymously during the Delphi. One study conducted a web-based anonymous survey, and a clinic based survey that was still anonymous. The final study conducted the Delphi completely anonymously during interviews.

Structure of the Delphi

Number of panels and rounds

The majority of studies included a single panel only (89%; 23/26 studies). In 13/23 studies the single panel comprised clinicians only. The remaining 10 studies comprised a single panel made up of mixed stakeholder groups. Three studies included two panels in their Delphi, all with separate panels for patients and clinicians.

The number of rounds in the Delphi studies ranged between one and six, with the majority including 2 or 3 rounds (73%; 19/26 studies). One of these studies included three rounds for clinicians and two for patients. The number of rounds was not clear in three studies. One study carried out a three round Delphi, but only round 3 was used for outcomes prioritisation. The remaining numbers of rounds conducted were one, four and six.

Administration of questionnaires

The method of administration of questionnaires is described in the Anonymity section above.

How was the original list of outcomes generated?

The general format of Delphi studies used to develop COS begins with identification of potential outcomes. It was not clearly reported how the initial list of outcomes was generated in five studies. In thirteen studies the initial list of outcomes was generated from a review of the literature; six of these combined this with something else including expert opinion, local stakeholders, discussion by the steering group or investigators, patient focus groups and a survey. One study collected information from Delphi participants prior to the Delphi, and another study presented existing COS recommendations and participants were asked to recommend others in the first round. In the remaining six studies the first round was used to generate the list of outcomes by asking participants to suggest outcomes. In two of these participants were asked to suggest outcomes within domains suggested by the steering committee.

Information provided to participants before the first round

The information provided to participants before the first round was not clear in 14 studies (54%). Twelve studies provided some description of the information provided to participants before the first round of the Delphi exercise. Eight of these were in relation to the study, including the process, the aims of the study and background information. Three studies provided a summary of the available literature. The final study gave both study information and evidence in advance of the Delphi.

A description of rounds

Table 8 in Appendix **3** 3 describes each round of each study, where this information was provided. How outcomes were kept in between rounds is also described in Table 8 (Chapter 3). This was not described in six studies, and the level of detail provided is variable in those that do report something about this.

A description of round 1 was provided in 22/26 studies (85%). One study was completed after one round of the Delphi, which although does not satisfy the criteria of a Delphi study, is included in this section because it reports to use this method. Of the 25 studies that had more than one round, 19/25 provided a description of round 2 (76%), 12/14 provided a description of round 3 (86%) and the two studies that had more than three rounds provided a description of subsequent rounds. Rounds 1-3 are summarised in Table 22.

Two studies were described as having more than three rounds. One study described having four rounds and in the final fourth round, the responses received from the third round were tabulated as the 'Draft gold standard.' This is reported as the fourth round in the report. The other study that reported more than three rounds included six rounds. In round 4, modified statements were presented requesting comments on their content and suitability for the proposal. They were then provided a draft paper for submission as a final proposal in Round 5 and asked to give final approval of the draft in Round 6.

Table 22: A summary of Delphi rounds 1-3

Round	n	Group score feedback	n
Round 1 (N=26)			
Outcomes rated using a 9 point numerical rating scale	2		
Outcomes rated using a 7 point numerical rating scale	1		
Outcomes rated using a 5 point numerical rating scale	4		
Outcomes rated using a 3 (patients) and 4 (clinicians) point numerical rating scale	1		
Participants asked to distribute 100 points among domains considered important	2		
Outcomes rated using a categorical rating scale (e.g. extremely important or not important)	3		
Create the initial list of outcomes	8		
Rate outcome domains (no further information)	1		
Not described	4		
Round 2 (N=25)			
Review group and individual scores from Round 1 and re-rate outcomes	7	Group mean and standard deviation	1
		Group median	1
		Group median, interquartile range, and total range (clinician panel); Group mean and total range (patient panel)	1
		Group median and interquartile range	2
		Not specified	2
Review the results from Round 1 and choose from them the outcomes that should be the core set	1	Overall group's percentage rating for each outcome	1
Review the most frequent responses from Round 1 and affirm agreement/disagreement	2		
First rating of outcomes (where the first round was outcomes identification)	6		
Continuing the list of outcomes (where the first round was outcomes identification)	2		
New items, re-worded items, and items for which there was disagreement and/or median rating of 4 (neither agreement nor disagreement) were re-rated in the second iteration.	1		
Not described	6		
Round 3 (N=14)			
Repeat of round 2	7	Mean and standard deviation	2
		Median	1
		Unclear	1
		Group median, interquartile range, and total range	1
		Median and interquartile range	2
Rank the top three most important outcomes	2	Group median and interquartile range	
Participants asked to distribute 100 points among domains considered important	1		
Asked to re-select the outcomes that should be part of the core set	1		
Agreement whether outcome should be in final set	1		
Not described	2		

How consensus was reached

The Delphi technique was the final method used to reach consensus in 15 studies (58%). Eight of these (53%) did not provide a definition of consensus, or description of how it was reached. Five studies used the proportion of participants recommending that an outcome should be included as a way of determining that consensus was reached. The required proportion of participants in four studies was 70%, and one study 66%. One study used a specific score to determine when consensus was reached for inclusion of an outcome. They used a scale of 1-7 and used a median score to determine whether an outcome should be included in the core set (see Taylor 2008 in Table 8 in Chapter 3). The final study that included a definition of consensus stated that if two of nine national stakeholders ranked it among the top 3 outcomes it was included.

Two other studies did not use a predefined level of consensus but provided some description: one simply stated that participants agreed on the overall recommendations at the end of the Delphi process, and the other stated that a steering group selected the ultimate core outcomes after the final round of the Delphi. In the latter, cluster analysis was performed prior to steering group discussion to further reduce the number of outcome domains and tools to make the number more practical for clinical trials. Outcome domains suggested by the group as being most important were classed as feasible or not.

The sequence of work was unclear in one study and so it is not clear whether Delphi was the final method to reach consensus. Furthermore, ten studies did not use the Delphi technique as the final method to reach consensus about outcomes, and went on to do additional work after the Delphi to determine the final core outcome set. In all but one study this further work was carried out face to face, including meeting(s), nominal group technique and workshop(s). In one study the recommendations underwent 'expert panel review' but it was not clear whether this was face to face. Eight of these studies (73%) did not provide a definition of consensus for the Delphi part of their study. Two studies used a proportion of participants recommending that an outcome should be included as a way of determining that consensus was reached, and these outcomes were then carried forward to the next stage of the work. The proportion of participants used was at least 60% in one study, and at least 75% in the other. One study stated that only outcomes with a median score of at least 6/9 on a Likert-type scale were considered in a subsequent consensus meeting, as consensus was not the aim of the Delphi.

4.4.3 Consensus meeting

As stated above, ten studies carried out additional work after the Delphi method to determine the final COS, and in nine of these it was clear that this was conducted using a face to face method. Thus, the following summary will relate to the nine studies that did further work on determining COS in a meeting(s) after the Delphi process.

Six studies held a single meeting, one study held two meetings (this study described the first meeting as exploratory, the second as consensus) and two held three meetings (in one study this was described as continuing the work in the first meeting, and in the second study as presenting and ratifying the core set); therefore there were a total of 14 meetings conducted relating to 9 COS. These meetings were the final method to reach consensus in all but one study where further work was done through post-meeting consultations among the group using emails, and the findings were made available on a website for public review. Five of the nine studies applied the Nominal Group Technique to their meeting. The level of reporting was variable, but no meeting was conducted the same way and they varied widely in duration from two and a half hours to four days. Table 23 shows a summary of the meeting details.

Structure of meeting(s)

Eight of the nine studies carried out a voting exercise, all of which took place after group discussion. The voting scale used was described in two studies; a six point scale for lists of statements from 1 (agree strongly) to 6 (disagree strongly) and a five point scale for lists of outcomes from 1 (being the least important) to 5 (being the most important). In the first meeting of the Mease study, priority ranking (first, second and third important domains to measure) was used. In the study's second and the third meeting, participants were presented with three choices to choose from for the presented outcomes.

Anonymity

Anonymous voting was conducted in four studies. An electronic voting system was reported as being used in the second meeting of the Gladman and the Schmitt study. Audience response methodology (description not provided) was used in the final meeting of the Mease study. The remaining studies did not report whether voting was anonymous.

Table 23: Meeting details

Reference	What happened during the meeting (in the order of the event)	Duration	Who led/assisted the meeting or the breakout groups
McGrath et al. (2008)	Discussion only	2 days	Meeting - senior author, two co-leaders of previous meeting
Dent, et al. (2008)	Anonymous voting that included discussion time	Not reported	Meeting - non-voting chairman
Schumacher et al. (2009)	A presentation, 2 breakout groups discussion, combined session (feedback the results of the discussion), voting at a plenary session	2 and half hours	Not reported
Gladman (2005)	Presentations, 3 breakout groups discussion (Nominal group process used, one group did anonymous voting), Combined session (feedback the results of the discussion)	Not reported	Each breakout group - a leader and a scribe
Gladman et al. (2005)	Presentations at a plenary session, 12 breakout groups discussion, combined session (feedback the results of the discussion), voting, presentation of the summary in the second plenary session	Not reported	Each breakout group had a scribe
Gladman et al. (2007)	A series of questions posed to audience, presentations, voting on the domains, breakout group discussion, voting to consider the domains into three categories	Not reported	Not reported
Mease et al. (2005)	Presentations, a panel discussion. voting (priority ranking), results presented in the plenary session	Not reported	Not reported
Mease et al. (2007)	Presentations, breakout group discussion, combined session (feedback the results of discussion), voting at the workshop and the plenary session	Not reported	Not reported
Mease et al. (2009)	Presentations, breakout group discussions, voting at the module and plenary session	Not reported	Not reported
Schmitt et al. (2010)	Presentations and discussions	3 hours	Not reported
Schmitt et al. (2012)	Presentations, 2 panels structured discussion (NGT applied), anonymous voting	2 days	Each group - one moderator and one rapporteur
Heiligenhaus (2012)	NGT process (discussion round, two ranking rounds)	2 days	Not reported
Douglas (2009)	NGT process (Discussion, voting)	Not reported	Not reported
Ruperto (2003)	NGT process Introductory lectures, three working group discussions, 5 NGT exercises carried out during the meeting including voting.	4 days	Three moderators with expertise in NGT

Participants

The number and type of meeting participants is described in Table 24. The number of participants ranged from 8 to 137, although the number of participants was not always reported (4/14; 29%). The type of participants was also inconsistently reported, with three studies not reporting this information, and a further five studies reporting who was included but not providing the number of participants per type. Of the 11 that did report who the participants were, 100% included clinical experts, and 46% (5 meetings) included patient representatives. In 8/14 meeting reports it was unclear whether the participants had taken part in the previous Delphi exercise; in the remaining six meetings, some or all of the participants had taken part in the Delphi.

How consensus was reached

Seven studies provided a definition of consensus. The level of agreement to be reached in all studies was expressed in terms of the proportion of participants recommending that an outcome should be included. This included 100% (n=1), 80% (n=2), 75% (n=1), 70% (n=2) and in the first and the third meeting of the Mease study at least 50% and 70% agreement was required. It was clear in three studies that the level of consensus had been determined a priori, it was not clear in the remaining studies.

Table 24: The number and type of meeting participants

Study	Method of identification of the sample of participants	Number of participants	Number (%) of clinical professionals	Number (%) of patient representatives	Other types of participants involved in the study (%)	Number of countries represented: Names (Continents)
McGrath et al. (2008)	People known to facilitator (same participants as the Delphi study)	26	17 (65)	0	FDA/NIH representatives n=5 (19) Industry representatives n=4 (16)	4: Canada, USA, UK, Sweden (North America, Europe)
Dent, et al. (2008)	People known to facilitator Published researchers (same participants as the Delphi study)	13 (including a non-voting chairman)	12 (92)	0	Clinical trial methodologist n=1 (8)	9: Australia, USA, Canada, Denmark, UK, Japan, Norway, Italy, Belgium (North America, Europe, Asia, Australasia)
Schumacher et al. (2009)	OMERACT conference participants (unclear if involved in Delphi)	77	Unclear who was involved	Unclear who was involved	Unclear who was involved	Not reported
Gladman (2005)	People known to facilitator (unclear if involved in Delphi)	Not reported	Not reported	Not reported	0	Not reported
Gladman et al. (2005)	OMERACT conference participants (unclear if involved in Delphi)	Not reported	Not reported	Not reported	Unclear if any other groups were involved	Not reported
Gladman et al. (2007)	OMERACT conference participants (unclear if involved in Delphi)	137	Unclear who was involved	Unclear who was involved	Unclear who was involved	Not reported
Mease et al. (2005)	OMERACT conference participants (unclear if involved in Delphi)	Not reported	Not reported	0	Academic and pharmaceutical based researchers n=not reported	Not reported
Mease et al. (2007)	OMERACT conference participants (unclear if involved in Delphi)	104	Not reported	Not reported	Unclear if any other groups were involved	Not reported
Mease et al. (2009)	OMERACT conference participants (unclear if involved in Delphi)	Not reported	Not reported	4 (unknown)	Researchers, statisticians, pharmaceutical industry representatives n=not reported	Not reported

Schmitt et al. (2010)	Conference participants, individuals with a known interest in eczema, and all those involved in Delphi exercise were invited to the meeting	Approximately 40	Unclear who was involved	Unclear who was involved	Unclear who was involved	Not reported
Schmitt et al. (2012)	People known to facilitator Published researchers (participants did include Delphi and 2010 meeting participants)	43	29 (67)	5 (12)	Methodologists n=5 (12) and pharmaceutical industry representatives n=1 (2)	10: UK, Netherlands, Germany, France, Sweden, Brazil, Israel, USA, Australia, Japan (Europe, South America, Asia, North America, Australia)
Heiligenhaus (2012)	Research working group (all took part in the Delphi)	14	14 (100)	0	0	9: Germany, UK, Switzerland, Finland, Spain, Netherlands, Denmark, France, USA (Europe, North America)
Douglas (2009)	Study steering committee members (all took part in the Delphi)	8	8 (100)	0	0	2: USA and UK
Ruperto (2003)	Clinical trial networks (unclear if involved in Delphi)	40	40 (100)	0	0	34: not reported

4.5 Discussion

A full discussion of the strengths and limitations of the systematic review is provided in Chapter 3.

COS developers have conducted different types of literature and systematic reviews as part of the process, with the majority undertaking a review of outcomes measured previously in clinical studies. Across studies that have done a systematic or literature review as part of their methods to develop a COS, the methodology differs considerably. Studies included different types of studies, had varying language restrictions, searched different date ranges and searched a variety of different databases, ranging between one and eight. Some went on to group outcomes, and studies that did so provided differing levels of detail about what was done. Furthermore, some studies excluded outcome categories such as cost and resource use that might be considered relevant to clinical trials. As highlighted in Chapter 3, it is currently unclear which of the multiple available frameworks are most suitable for categorising outcomes. Having an agreed framework for categorising outcomes might aid COS developers' in doing so. This issue is discussed further in Chapter 8.

Main features were often missing from reports. A third or more of studies did not report the type of studies that were included (39%), did not provide information about language restrictions applied (44%), the date ranges searched (37%) or the databases searched (37%). Furthermore, 30% did not report the results or main findings from the review. Further work could be done to identify the key features that should be reported by all studies including a literature/systematic review as part of the COS development process. These results also highlight the need for methodological guidance on how to design such a systematic review.

The Delphi technique is another method that has been used in different ways in COS development. It was the final method for reaching consensus for just over half of the studies that used this method (58%), but others went on to do further work. Where the Delphi technique had been used, it had been done variably; across studies the composition of the groups was variable and the process was never the same. Less than half of the studies included participants outside of the EU and US, and half of the studies included participants from one country only (UK, US and Italy). Patient participation was relatively low, and their involvement in the process differed. Furthermore, the identification of one

study that only included one round of Delphi raises the question whether this study should be categorised as such. This highlights the need for a way of quality assessing COS studies and the methods they purport to use.

The quality of reporting of key aspects as identified by Sinha et al [59] was variable. As this checklist was published in January 2011, I separated studies to look at reporting before and after the publication of this checklist. It would seem that there have been some improvements in the reporting of some of the recommended characteristics. However, other key characteristics seem to be still poorly reported, including 'Information provided to participants before the first round', 'How the overall group response was fed back to participants' and 'A priori definition of consensus about whether an outcome should be measured.' Although some studies did provide a definition of consensus, it was often unclear whether this was a priori. Reporting of Delphi studies seems improved but there are many key features still not being consistently reported.

Where a consensus meeting was used as the final method to reach consensus, no meeting was conducted in the same way. They varied in duration, how many were held and what they did at each meeting. The duration of meetings was often not reported, as was who led the meeting and who was involved. What happened at meetings was described in varying levels of detail. Most meetings included a voting exercise, although the method of voting or scale used was often not reported. Anonymity was described in approximately half of the studies. A definition for consensus was provided in the majority (78%) although it was only clear that it was an a priori definition in three of the seven studies.

4.5.1 Future work

The uptake of COS depends on the user's ability to be able to assess the core set for suitability of use. This requires an assessment to be made on the appropriateness of the methods used as well as the reporting of the key features of these methods. Poor reporting makes it hard to appraise, thus key features need to be reported.

This chapter further exemplifies the variability in COS methodology. Furthermore, these findings corroborate the conclusion of the main systematic review (reported in Chapter 3) that there is a need to improve the standards of reporting in these studies. As discussed in Chapter 3, a generic COS reporting guideline is currently being developed, but here I

highlight the need for reporting guidelines specific to each type of study. The checklist developed by Sinha et al specifically for Delphi studies seems to have gone some way in helping to improve the reporting of Delphi studies. However, key characteristics were sometimes still missing from Delphi study reports so there is still a need to help improve this. One possible solution could be the promotion of these guidelines through the COMET Initiative to make sure COS developers are aware of them. It is not known whether more recent studies have used the Delphi checklist to report their studies, and indeed whether they have found this helpful.

4.5.2 Summary

Given the variations in methodology between studies, it would be helpful if we understood developers' rationales behind their methodological decisions. There is a need to determine how best to develop COS. This chapter provides more detailed information about the methods that we know are being used presently to develop core sets. By exploring choices of methodology we will better understand why developers have done things so far, enabling us to better inform how COS developers might do this work in the future. The work in this chapter affirms the need for further work to explore these issues, which is the focus of Chapters 5 to 7.

Chapter 5: A survey of core outcome set developers

5.1 Background

The systematic review of COS identified 250 reports relating to 198 studies. The review showed that a range of methods have been used, and applied in a variety of ways, to develop COS. Furthermore, of the 178 studies that described the methods they used to determine the COS, 164 (92%) did not provide an explanation regarding their choice of methodology. The reasons behind the methodological choices in the 14 studies that did provide explanation are provided in Table 7 in Chapter 3. The most common reason provided was that the methods had been used previously to develop COS or were well-recognised methods for eliciting expert consensus to inform guidelines.

To my knowledge, there is little guidance about how to conduct or report COS studies and it is currently uncertain which of the various methods reported are the most suitable, feasible and efficient. The methods used in the development process may have an impact on the conclusions derived. Consensus work undertaken by three different groups in the same clinical area (paediatric asthma) [103-105], but using different methodological approaches and involving different stakeholder groups, resulted in inconsistent outcomes being rated as important; although there was some overlap in the outcomes chosen. Research to investigate COS developers' choice of approach may help to illuminate the reasons for these differences, as well as highlight areas of COS development that would benefit from improved guidelines and recommendations. It is important to investigate COS developers' choice of approach as this is a new area of research, and we need to understand more about why these COS developers chose the methods that they did.

The COMET Initiative Management Group published some preliminary suggestions around issues to consider when developing a COS [38]; these included steps to encourage the implementation of COS recommendations once they have been developed. To increase COS uptake, it is recommended that developers consider engagement with the relevant Cochrane Review Groups, clinical guideline developers, research funders, journal editors, regulators such as research ethics committees, and trial registries. A recent series of

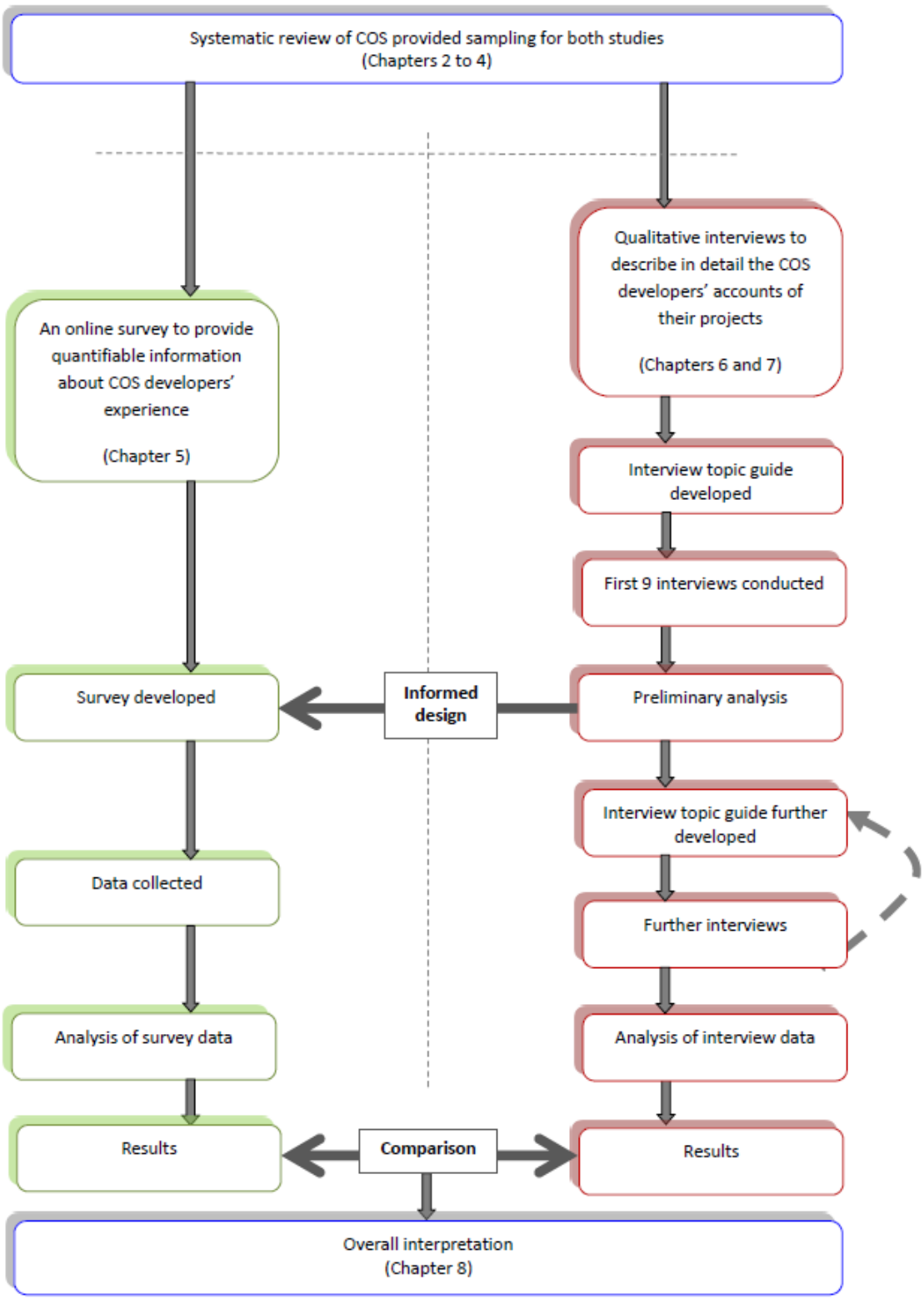
papers in the Lancet addressing waste in research also highlighted that efforts to improve the quality and usefulness of health research will have a greater chance of success if research funders and journals more actively endorse and implement these initiatives [37]. The systematic review of COS (discussed in Chapter 3 and 4) showed that only two papers explicitly suggested or discussed plans for the implementation of their recommendations. Giacoia's report [80], which is linked to three COS studies, stated that their proposed recommendations will be considered as a requirement by a governmental agency (NICH); and Cranney [81] stated that the proposed core measures were being submitted to key international research groups in osteoporosis to promote their acceptance and implementation. Implementation may not have been a priority for the majority of COS developers or it may have been beyond the scope of these papers; it is therefore important to explore this further.

5.2 A mixed methods approach

This thesis presents a mixed methods approach to explore COS development. Mixed methods research is where the researcher(s) combine elements of qualitative and quantitative research approaches for the broad purposes of breadth and depth of understanding and corroboration [106]. I drew on qualitative methods (interviews) and an online web-based survey to provide the first comprehensive account of COS development. As an emerging area of research where little is known about COS development processes, and developers' choices of methods in particular, mixed methods allowed me to quantify the current situation and COS developers' experience, and at the same time explore the factors and actions that shaped those experiences [107]. In particular, this study used a Triangulation Design to obtain different but complementary data on the same topic for comprehensiveness [108]. Data was collected in parallel and received equal status in this study, and although these data were initially analysed separately, the interpretation of the findings were integrated as indicated in the final Discussion chapter (Chapter 8). The study design is depicted in a visual model in Figure 5.

The qualitative work (interviews) aimed to examine researchers' accounts and opinions of their work to develop COS, with the goal of identifying potentially important but previously unanticipated issues; this work is described separately in chapters 6 (methods) and 7 (results). This chapter will outline the web based survey, the content of which was informed by the first 9 interviews conducted.

Figure 5: Visual model of mixed methods design



5.3 Aim

To provide quantifiable information about COS developers' experience; to explore similarities and differences in those experiences, and to further understand the COS development process.

5.4 Methodological approach

The survey was constructed using SelectSurvey.NET (http://www.liv.ac.uk/csd/survey/index.htm), an online survey package provided by the University of Liverpool. The benefits of using this software included the facility to incorporate filter questions (whereby depending on the responses, questions are automatically skipped to the next appropriate question), the survey can be programmed individually for the study purposes, and the answers can be automatically programmed to download into a database [107]; the latter was particularly important because the survey asked COS developers to clarify the data that had been previously extracted from their paper as part of the systematic review, or provide additional information if this was missing. As such, SelectSurvey.NET was deemed more flexible to fit the purpose of the survey than using existing online survey software such as SurveyMonkey or SurveyGizmo.

The survey was sent as a link within a personalised email inviting prospective respondents to visit a website where the survey could be found and completed. Adopting a personalised approach has been suggested to increase odds of response [109]. An example of the invite email can be found in Appendix 4. Odds of response have been shown to increase by more than a quarter when follow-up contact is made with those who do not respond to the initial survey [109], therefore two further mailings were sent to non-responders, two and four weeks after initial emails. Those who had not responded six weeks after the date of initial emails were recorded as non-responders overall. Other strategies suggested to increase response were utilised, including not using the word 'survey' in the subject line of the email, using a white background for the survey, including a deadline to reply, and keeping the survey short.

5.4.1 Structure of the survey

COS developers were asked to answer a few short questions about their COS work, including: how the COS study came about; how they decided which methods to use, and

whether implementation was a priority for them/their group. COS developers were asked to clarify or confirm the data that has been extracted from their paper as part of the systematic review (or provide additional information if missing), including: target population age; interventions covered; and methods used to develop the COS, including the stakeholders included in the process. Each COS developer received a slightly different version of the survey depending on the information required. The survey can be found in Appendix 5.

5.4.2 Participants

The contact author of the COS publications identified in the systematic review were contacted and asked to participate in the short survey. Where an author was involved with multiple COS, they were asked to complete the survey for the most recent COS only. COS authors who had already taken part in the qualitative study (interviews) were excluded.

5.4.3 Analysis and presentation of results

Data is described in narrative form, and the findings provided in text and table format with illustrative comments from responses to open questions. Responses to open questions were categorised into common topics. First, free text data was read and re-read several times to gain a broad understanding of the topics covered. Recurring topics were identified, and the data was labelled. Data with the same labels were brought together and grouped into common themes. Any discrepancies were resolved though discussion with my supervisors (PW and BY).

The survey response and completion rates were calculated. The response rate refers to the number of people who initiated a response to the survey divided by the total number that the survey was sent out to. The completion rate refers to the number of surveys filled out and submitted. This is calculated by dividing the number of surveys filled out by the number of surveys started by respondents.

Differences between responders and non-responders were examined by comparing the year of study and whether patients were involved in the development of the COS. These factors were examined because it was felt that they might be associated with whether a response was received or not.

5.5 Results

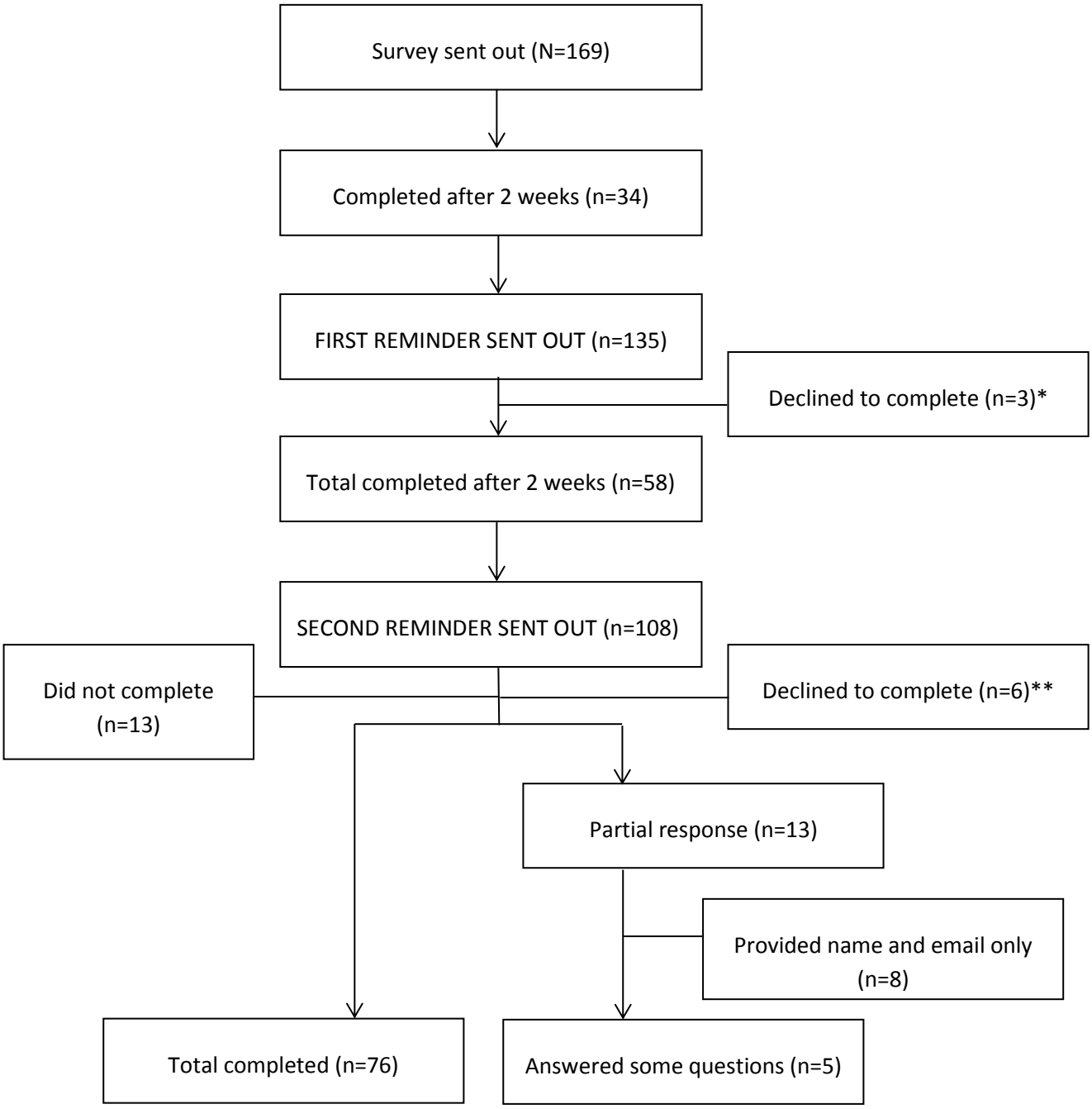
One-hundred and ninety eight COS studies were identified in the review. Twenty nine studies were excluded for the following reasons:

- already interviewed (n=12)
- invited to interview awaiting response(n=3)
- accepted interview invite, waiting to be scheduled (n=2)
- where there were duplicates (where the lead/contact author was the same for multiple COS) we asked them to complete the survey for the most recently published COS only (n=7)
- author was deceased (n=1)
- unable to find active/correct email address (n=4)

The survey was therefore sent out to 169 COS developers. The study flowchart is shown in Figure 6. Four versions of the survey were set up, and each COS assigned a version depending on what information we needed about their COS development methods and participants.

The survey had a response rate of 53% (89 of 169 invitees), and a completion rate of 85% (76 of 89 started). Eight of the thirteen partial responders only provided their name and email address, so did not provide any useful information. The remaining five answered some questions, meaning that there were a total of 81 responses to the survey. From this point forward these will be classed as responders (n=81), with all other responses classed as non-responders (n=88).

Figure 6: Flowchart showing participants in the survey



**One author responded to say it was too long ago and they could not remember enough details to complete the survey, one author had retired (co-author deceased), and the third made a referral to their co-author but this individual had already been contacted about the survey in relation to another study.*

***Unable to complete due to other commitments, sent to co-authors but no response.*

5.5.1 Response bias

I looked for differences between responders and non-responders (Table 25). The year of publication of the earliest identified report for each study was used for comparison and showed that a higher percentage of responses were from 2010-2013 studies. COS developers were also more likely to respond if patients had been involved than not.

Table 25: Differences between responders (n=81) and non-responders (n=88)

Year of study			
	Total	**Responders n (%)**	**Non-responders n (%)**
1980-1984	1	1 (1.2%)	0
1985-1989	1	1 (1.2%)	0
1990-1994	13	6 (7.4%)	7 (8%)
1995-1999	25	7 (8.6%)	18 (20.5%)
2000-2004	30	15 (18.5%)	15 (17%)
2005-2009	55	22 (27.2%)	33 (37.5%)
2010-2013	44	29 (35.8%)	15 (17%)
Patients involved			
Yes	24	15 (18.5%)	9 (10.2%)

5.5.2 Sample characteristics

The sample characteristics for the 81 respondent COS are presented in Table 26. Surveyed COS covered a broad range of disease categories, representing all those identified in the systematic review except endocrine and metabolic conditions, and wounds. Surveyed COS also covered a broad range of population and intervention characteristics and reflected all of the categories identified in the systematic review of COS.

Table 26: Sample characteristics of 81 surveyed COS

	n (%)
Population characteristics	
Both adults and children	10 (12)
Adults	57 (70)
Children (including neonates)	10 (12)
Older adults only	4 (5)
Intervention characteristics	
All intervention types	26 (32)
Drug treatments	41 (51)
Drug only	*26*
Drug and devices	*2*
Drug and behavioural	*5*
Drug and procedure	*1*
Drug and surgery	*2*
Drug and device and behavioural	*1*
Drug and procedure and device	*1*
Drug and procedure and device and surgery	*2*
Drug and procedure and surgery	*1*
Surgery	5 (6)
Surgery only	*2*
Surgery and procedure and device	*2*
Surgery and procedure and device and behavioural	1
Procedure	5 (6)
Procedure only	*1*
Procedure and device	*2*
Procedure and behavioural	*1*
Procedure and behavioural and device	*1*
Devices	1 (1)
Other	3 (4)
Vocational rehabilitation	*1*
Models of maternity care	*1*
Screening	*1*
Disease category	
Neurology	12 (15)
Rheumatology	12 (15)
Cancer	11 (14)
Heart & circulation	9 (11)
Orthopaedics & trauma	7 (9)
Infectious disease	4 (5)
Gynaecology	3 (4)
Urology	3 (4)
Dentistry & oral health	2 (3)
Kidney disease	2 (3)

Lungs & airways	2 (3)
Mental Health	2 (3)
Tobacco, drugs, & alcohol dependence	2 (3)
Anaesthesia & pain control	1 (1)
Blood disorders	1 (1)
Ear, nose & throat	1 (1)
Gastroenterology	1 (1)
Genetic disorders	1 (1)
Health care of older people	1 (1)
Intensive care	1 (1)
Neonatal care	1 (1)
Pregnancy & child birth	1 (1)
Skin	1 (1)
Methods used	
Semi-structured group discussion only	25 (31)
Workshop	*7*
Meeting (meeting, colloquium, conference where not described as consensus development conference)	*17*
Round table discussion	*1*
Unstructured group discussion only Descriptions include task force, work group, working group/party, committee, board, panel	6 (7)
Consensus development conference only	6 (7)
Literature/systematic review only	6 (7)
Delphi only	2 (3)
Survey only	2 (3)
Mixed methods (see descriptions below)	34 (42)
Delphi + another method(s)	7
NGT	*2*
NGT + literature/systematic review	*1*
Literature/systematic review	*3*
Literature/systematic review + consensus conference	*1*
Semi-structured group discussion (listed which method) + another method(s)	17
Workshop + literature/systematic review	*1*
Meeting + literature/systematic review	*8*
Workshop and meeting	*2*
Workshop/meetings + web-based consultation	*1*
Workshop + survey + literature/systematic review	*3*
Round table discussion + literature/systematic review	*1*
Meeting + focus group(s) + survey	*1*
Consensus development conference + another method(s) + Literature/systematic review	1

Unstructured group discussion + Literature/systematic review	3
NGT + another method(s)	3
Survey + interview	*1*
Semi-structured discussion (workshop & meetings)	*1*
Workshop + literature/systematic review	*1*
Literature/systematic review, public presentation and debate	2
Literature/systematic review, survey and open discussion	1
Participant groups involved	
Clinical experts + Research experts (only or Authorities/Providers/Others)	30 (37)
Clinical expert only	23 (28)
Clinical experts + patients (only or Research experts/Authorities/Providers/Others)	20 (25)
Clinical experts + Authorities (only or Providers/Others)	5 (6)
Clinical experts + Others	2 (3)
Patients only	1 (1)

5.5.3 COS motivation

How did your core outcome set study come about?

We asked 'How did your come outcome set study come about?' Response options were fixed, with a free text option to provide more information for those indicated with a *, and multiple responses were allowed. All 81 respondents answered this question. The motivations for COS development are shown in Table 27. The most frequent answers indicated that the drivers for the COS was the heterogeneity in what outcomes were being measured or reported in trials/research, or heterogeneity in the way outcomes were being measured, so studies were measuring the same outcomes using different tools/instruments/measures. COS developers indicated that there was not an existing COS that could be used. Nine COS developers responded that a COS did exist but they did not think it was good enough or suitable for use.

Table 27: Motivation for core outcome set development

Response item	n	Illustrative comments
There was heterogeneity in which outcomes were being measured (studies measuring/reporting different outcomes) in trials/research	52	*There are a variety of outcomes tools, clinical measurement tools that are used for measuring [disease name]. (S62)*
There was heterogeneity in the way outcomes were being measured (studies measuring the same outcomes using different tools/instruments/measures) in trials/research	52	*A variety of core OMs existed but their use in prehospital/EMS [Emergency Medical Services] research was limited or heterogeneous and few had been validated. (S42)* *There was significant heterogeneity in the definitions and measurement of outcomes in earlier studies. (S57)*
I/we thought there was something missing in the outcomes being measured/reported in research	46	*Some important outcomes were not reported in earlier studies. We aimed to encourage standardisation. (S57)*
There was no existing core outcome set that we could use for our study	40	
I/we thought there was something missing in the outcomes being measured/reported in clinical practice	28	*Additionally, although there is a standard of intervention care, the treatment protocols are often deviated from based on individual needs. (S62)*
There were outcomes being measured but not reported in trials/research	16	
It was part of a research prioritisation study*	13	*Prior to our trial there were no multi-site, combined-treatment randomized clinical trials Class 1 studies for treatments of this disorder. (S39)*
The outcomes that were being measured in research were not applicable/relevant to clinical practice	13	
I/we were conducting other research (e.g. a systematic review or a trial)*	11	*We lead a clinical trials network and needed a standardized outcome for the interventional trials in [disease name]. (S21)* *We were conducting an observational trial and were discussing what the gold standard for success*

		should be for our subject population. We could not agree on what definitely needed to be included, therefore we chose to survey experts in surgery and rehab. (S23) *A Cochrane review of [disease name}. (S61)* *Conducting trials of potential medications to treat [disease name]; FDA guidance was a draft form and did not meet the needs of investigators or industry. (S79)*
A core outcome set existed but we did not think it was good enough/suitable*	9	*2 core sets were available, but they were both too broad. (S14)* *It had been identified that the patient voice was not included when the core sets had previously been developed and therefore there were some outcomes important to patients that were not included. (S44)* *Previous proposals were not detailed enough. (S45)*
I/we were motivated by work that had been done in another speciality	8	
Other reasons provided*	5	*I was requested to write the paper. (S11)* *We were developing guidelines for the American [disease area] Association to use in reviewing products for awarding the ADA Seal, and possibly for the FDA to use in similarly assessing products for treatment of [disease name]. We were not explicitly developing a core outcome set.(S37)* *Well established set of outcomes in the field, but no large trials on the specific topic. (S41)* *A commissioned study. (S49)* *The society wanted a reference document. (S63)*

5.5.4 COS methods

How did you decide on the methods you used?

Table 26 shows the methods used to develop COS for the 81 studies included in the survey. The most frequent method was semi-structured group discussion (n=25; 31%), followed by mixed methods using semi-structured group discussion in combination with another method, most commonly a review of the literature. The consensus development conference, Delphi technique and literature reviews were also used.

We asked COS developers "How did you decide on the methods you used?" Response options were fixed, with a free text option to provide more information for those indicated with a *, and multiple responses were allowed. All 81 respondents answered this question and the results are presented in Table 28.

Most respondents reported that they had based their choice of methods on the literature (previous work), specifically citing the OMERACT group's work for rheumatoid arthritis and the HOME group for eczema as examples of previous work that had informed their COS development methods. Respondents also reported that expert advice informed their choice of methods, with 6/51 (12%) of respondents who cited this giving this answer exclusively. COS developers also reported that the available resources influenced the methods that were used; comments related to funding and personnel available to carry out the work. The author of one recent study also referred to the COMET website as an available resource that had helped in deciding on the methods to use.

Table 28: Rationale for choice of methodology

Response item	n	Illustrative comments
Based on the literature (previous work)*	59	*OMERACT, HOME for eczema, COMET website.(S36)* *We were aware of previous reports and the variability of reporting in them.(S38)* *We used Nick Black's publication on consensus methods, some common sense, and developed a framework for doing systematic reviews of outcome measurement properties. (S61)*
Based on expert advice*	51	*Informally we sought advice from experts. (S38)* *Expert panel of individuals knowledgeable of prior work.(S43)* *[Name] was very helpful in advising and participating. (S70)*
Own experience with same methods before for core outcome set development	31†	*Many of us had done similar things before and we had applied for and received funding from [funding body] to do this in the light of recent trials. (S45)* *We used a workshop with 20-22 experts in different fields of the disease. This was easy to organise and we had previous experience with it.(S48)*
Suited our situation and circumstances*	27	*There was a limited group of researchers involved in the study. (S55)* *Since an international meeting was held, we tried to make use of this opportunity and held a workshop on outcome measurement and patient's assessment. (S65)*
Based on the resources available*	27	*Surveys of professionals has been used previously in [disease area], was relatively low cost, and was achievable with the work of mostly one person. (S23)* *OMERACT, HOME for eczema, COMET website.(S36)*
Problems with other methods*	15	*No illustrative comments*

† *Disease areas relate to: Dentistry & oral health (n=1); Health care of older people (n=1); Intensive care (n=1)*
Kidney disease (n=1); Mental health (n=1); Lungs & airways (n=2); Heart & circulation (n=2); Infectious disease (n=2); Urology (n=2); Neurology (n=5); Cancer (n=6); Rheumatology (n=7)

How did you decide who to include as participants in your core outcome set development work?

Table 26 shows the participant groups included in the development of the COS for the 81 studies included in the survey. We asked "How did you decide who to include as participants in your core outcome set development work?" Response options were fixed, with a free text option to provide more information for those indicated with a *, and multiple responses were allowed. All 81 respondents answered this question and the results are presented in Table 29, cross-classified by the stakeholder groups involved. The most frequent answer reported by respondents was to involve stakeholder groups with experience of or knowledge about trials or research. Experience with clinical practice was the second most frequent answer. Wanting an international perspective was also a frequent answer.

Experience of living with/caring for someone with a condition and being able to see things from the patient perspective were highest in those that had patient participation in the study. However, half of the COS developers who gave one of these answers did not include patients in the process.

Table 29: Rationale for choice of stakeholders

Response item	**Stakeholder groups involved**				
	Clinical expert only *n=23*	Clinical expert + other (non-patient) *n=37*	Clinical experts + patients (only or other) *n=20*	Patients only *n=1*	**Total**
Experience with/knowledgeable about trials/research	19	34	17	0	**70**
Experience with/knowledgeable about clinical practice	16	30	15	0	**61**
We wanted an international perspective	19	24	14	0	**57**
Experience of people who are involved in decision-making about treatment	9	17	9	0	**35**
To represent a broad view	7	19	8	0	**34**
To help with implementation and uptake later on	6	9	5	0	**20**
Experience of living with/having/caring for someone with the condition	4	4	8	0	**16**
Able to see things from the patient perspective	2	6	7	1	**16**
We wanted a national perspective	3	6	5	0	**14**
We wanted a local perspective	0	1	1	0	**2**
Other	1	1	0	0	**2**
Total	**86**	**151**	**89**	**1**	**327**

5.5.5 Differences in the outcomes thought to be important

We asked "Were there any differences in the outcomes thought to be important by different stakeholder groups?" All 81 respondents answered this question. 24/81 COS developers did not look for this; 18 of these did include more than one stakeholder group in the process. Twelve COS developers responded that it was not relevant to them as they only included one group of stakeholders in the process; however, a cross check with the stakeholders included indicated that 5/12 of these had involved more than one group. These were typically clinical and research experts which might suggest that COS developers saw these as sufficiently similar to combine as one group, and did not differentiate these as different stakeholders. Of the remaining 45, twenty-six indicated that there were no differences, but on further checking it would seem that four of these only included one stakeholder group. Nineteen indicated that there were differences in the outcomes thought to be important by different stakeholders. Only one of the 19 who reported differences in the outcomes thought to be important by different stakeholder groups discussed these differences in their publication.

A range of differences were reported. Three COS developers did not specify what those differences were; two COS developers described differences in relation to 'how' outcomes were measured and one described international differences but not specific to particular stakeholder groups:

> *Because of international variation in available technical and clinical resources for clinical trials, there were international variations in preferences for main outcome measures. (S57)*

2/19 described differences between clinical professionals from different parts of the world:

> *Different opinions between US and European professionals due to differences in practice and reimbursement. (S1)*

4/19 described differences between clinical professionals when multidisciplinary groups were included:

> *The differences were related to the profession and the interventions; audiologist preferred audiological measurements, psychotherapists QoL measurements. (S65)*

2/19 described differences between clinical professionals and researchers:

> *There are different measures noted as clinically relevant that are difficult to capture in the context of a research trial, much discussion centered on around how to best objectify these variables. Differences were primarily between the clinical researchers and clinical experts.* (S62)

2/19 described differences between the FDA and another stakeholder group:

> *FDA favored clinically meaningful outcome measures, while industry favored MRI outcomes. (S24)*

Of the 45 respondents who answered yes or no to the question about whether there were differences in the outcomes thought to be important by different stakeholder groups, seven involved patients as well as health professionals in the COS development process. Of the seven studies three reported differences in the outcomes deemed important by these groups:

> *Patients have other priorities than physicians, e.g. function higher than cosmetics. (S26)*

> *Patients place more emphasis on pain. Healthcare professionals place more emphasis on controlling inflammation. (S60)*

5.5.6 COS developer experiences of the COS development process

What do you think were the main strengths of your study?

Respondents were asked to list up to five strengths. 76/81 respondents answered this question. Respondents provided a range of 1 to 5 strengths; the average (mode) was 3. A total of 211 strengths were listed and classified, and a total of 237 categories were assigned as some involved more than one category. For example, "*lots of patients (compared to previous core outcomes projects)*" was provided as one strength and therefore included as 1/211 strengths; it was categorised as both 'Stakeholder involvement' and 'Sample size' so included in 2/237 categories. A list of classifications can be found in Appendix 6, and the frequency of categories is shown in Table 30.

Table 30: COS developers' perceived strengths of their COS studies

Domain	Category	Frequency
Stakeholders (n=66)	Stakeholder involvement	57
	Sample size	7
	Patient research partners	1
	Response rate	1
Consensus (n=60)	International	25
	Consensus	15
	Standardisation	8
	Consensus definition/parameters	2
	Evidence based	7
	Experience based	3
Methods (n=45)	Process	27
	Literature review	12
	Method used	3
	Outcomes framework	2
	Scoring system	1
Design (n=42)	Wider trial design	13
	How to measure	13
	Novel	11
	Scope	2
	Starting point	2
	Resources	1
Implementation (n=22)	Endorsement	5
	Relevance to practice	5
	Relevance to research	5
	Impact	2
	Uptake	2
	Dissemination	1
	Relevance to regulators	1
	Prospective validation	1
Other (n=2)	Timely	2

Stakeholder involvement was the most frequently listed strength by COS developers. Comments included involving a '*broad group of experts and professionals*' (S49), '*involvement of multiple stakeholders – regulatory, industry, academics, patient societies*' (S24), as well as stakeholders with different, or '*large experience in the particular field*' (S66).

COS developers also described including key leaders in their field. Patients being involved as participants was also listed as a strength by some COS developers. One COS developer also described the involvement of patient research partners '*at every stage of the study*' as a strength of their study (S44). The involvement of industry representatives '*as full participants from the start*' was also noted as a strength of one COS developer's study (S25).

Another frequently listed strength was that the COS was international. This included both international representation in the process and achieving international consensus. Adopting a consensus process was also seen as another strength of these studies. Interestingly, some COS developers reported that this work was novel in the field and this was also seen as a particular strength. Two COS developers also perceived the work being *'timely'* as a strength *(S24, S42)*.

COS developers described endorsement of their work as one of its strengths, including 'p*otential for regulatory endorsement' (S16) and 'international representation and buy-in' (S18).* Uptake of the core outcome set was also perceived as a strength:

> *Used in subsequent clinical trials, allowing meta-analyses. (S48)*
>
> *Widely adopted. (S60)*

Others wrote about impact in a more general way:

> *Had a significant impact on future outcomes research in [disease area]. (S53)*
>
> *It has had an enduring impact. (S61)*

COS developers described an overall relevance to research, but also to '*clinical practice, trials, [and] regulatory agencies' (S6).*

What do you think were the main challenges you experienced over the course of your study?

Respondents were asked to list up to five challenges. 75/81 respondents answered this question. 5/75 stated that they did not experience any challenges in doing this work. Other respondents provided a range of 1 to 5 challenges; the average (mode) was 1. A total of

121 challenges were listed (137 categories assigned). A list of classifications can be found in Appendix 7, and the frequencies of categories are shown in Table 31.

Table 31: Challenges experienced over the course of COS development

Domain	Category	Frequency
Consensus (n=47)	Lack of data to support recommendations	14
	Achieving consensus	12
	Differences in opinion	7
	Different interests	3
	Expert biases	3
	Conflict of interest	2
	Opinion vs data	2
	Prioritising outcomes	2
	Including all views	1
	Multiple domains important	1
Design (n=36)	Resources	14
	How to measure	11
	Wider trial design	4
	No experience/knowledge of COS development	3
	Novel	2
	Feasible vs important	2
Methods (n=20)	Process	18
	No guidance	1
	Finding relevant articles	1
Stakeholders (n=17)	Patient involvement	5
	Participants	3
	Size of the group	3
	Language [international]	2
	Response rate	1
	Accessing participants	1
	Participant burden	1
	Recruitment	1
Implementation (n=14)	Implementation	3
	Changing practice	3
	Validation of outcomes	3
	Keeping up to date	2
	Publishing	1
	Dissemination	1
	Generalisability of COS	1
Other (n=3)	General†	2
	Poor understanding of disease	1

† *Heterogeneity. (S63)*
Lack of gold standard outcome. (S45)

Achieving consensus was described as challenging by COS developers, and was described as the main challenge by one COS developer (S54). This theme was echoed in another challenge described by COS developers, that is differences in opinions and interests between experts. One COS developer described the challenge of *'smoothing political differences,'* (S18) and another wrote that *'experts are disinterested in others views'* (S60). Having a lack of data was seen as another challenge experienced by COS developers, and this was described as 'finding *the balance between expert opinion and data' (S25) with 'limited literature to support recommendation' (S64).*

Respondents saw avoiding the potential for bias and managing conflicts of interest as further challenges for COS developers, sometimes in relation to the measurement instruments that participants' had previously developed.

> *To deal with the different (hidden) agenda's/motivations of the professionals. (S1)*

> *Some participants insisted in the instruments they developed and used. (S65)*

COS developers described experiencing challenges associated with resources, commonly funding, time or technology:

> *Financial limitations for support of experts. (S47)*

> *Time required to complete the work. (S61)*

> *Managing Delphi process through ICT. (S7)*

One COS developer went further to write that '*Speed meant we made a few mistakes,' (S61).*

COS developers listed many challenges in relation to the COS development process itself. Challenges included uncertainties about how to structure the process, organising stakeholders, difficulties with arranging meetings and keeping participants motivated. One part of the process described as challenging was narrowing down the long list of outcomes:

> *Narrowing down the potential research outcomes to a small enough list.* (S22)

> *Making the work relevant- we worked hard to hone the list to a manageable realistic level. (S62)*

Challenges associated with involving patients in the process were listed. Comments about the challenges involved included:

> *How to label the outcomes important to patients - to use their own language or map onto existing outcomes? (S44)*

> *Explain outcomes and why this is important to non-clinical participants (patients). (S36)*

Implementation and dissemination was also described as a challenge. One COS developer described tensions with other professional bodies:

> *The American Academy of [disease area] felt that our group, and the American [disease area] Association, were usurping its prerogative. (S37)*

Furthermore, implementation was seen as challenging because '*changing practice is difficult*,' (S60), and '*old methods [are] ingrained*' (S6).

A lack of methodological guidance (S7), lack of knowledge (S40), and lack of experience of COS development (S25, S48) were also seen as challenges. The relative novelty of COS development was also described as a challenge:

> *First of its kind, so everything was 'new', many things developed 'on the fly'. (S25)*

What do you think were the main limitations of your study?

Respondents were asked to list up to five limitations. 78/81 respondents answered this question. 4/78 reported that they did not think there were any limitations to their study. Other respondents provided a range of 1 to 5 limitations; the average (mode) was 2. A total of 123 limitations were listed. A list of classifications can be found in Appendix 8, and the frequencies of categories are shown in Table 32.

Table 32: COS developer perceived limitations

Domain	Limitation	Frequency
Consensus (n=34)	Lack of data to support recommendations	27
	Theoretical	3
	Expert bias	2
	Inability to recommend a COS	1
	Lack of preceding consensus	1
Stakeholders (n=30)	Stakeholder involvement	21
	Number of participants	6
	Language difficulties	2
	Response rate	1
Implementation (n=21)	Keeping up to date	7
	Implementation	7
	Endorsement	2
	Intended use (Trials vs practice)	2
	Lack of validation of COS	2
	Non-specific results	1
Other (n=20)	How to measure	13
	Other†	5
	Size of COS	1
	Limited knowledge	1
Methods (n=10)	Process	8
	No gold standard	2
Scope (n=8)	Disease/population	3
	Narrow focus	3
	Study design	1
	Intervention	1

† *Complexity of the situations to apply the methodology. (S50)*
It was not a study. (S53)
Over-emphasize drug effects. (S60)
Specific agendas by some experts not included in this chapter. (S79)
Differences between centres. (S41)

The most frequently listed limitation was a lack of data on which to base recommendations:

> *Evidence base lacking for some decisions. (S34)*

> *Limited evidence behind recommendations. (S64)*

COS developers felt that basing their recommendations on expert opinion in the absence of data was one of the main limitations of their work:

More consensus-based than evidence-based. (S27)

At the end it is expert opinion, not evidence. (S28)

Conclusions based more on expert view than trial data. (S45)

Similarly, COS developers wrote about COS development as being theoretical or conceptual as another limitation.

Lack of stakeholder involvement was often listed as the main limitation. COS developers most commonly described a lack of patient involvement. Other stakeholder groups listed as absent from their study included regulators and industry representatives. Two COS developers listed '*expert bias*' as another limitation (S42 and S63). Others made more general comments about limitations of stakeholder involvement, including a lack of international perspectives and '*only professionals [being] involved' (S74).* The number of participants included and the representativeness of included samples was seen as another limitation:

Too few cardiologists. (S28)

Although lots of patients replied to our project, their number was still lower than clinicians, hence possibly they were underrepresented. (S36)

Uptake, implementation and endorsement were other limitations listed by COS developers:

The work was sound, but the American Academy of [disease area] felt compelled to in essence redo and overwrite it with its own document. (S37)

That despite firm recommendation, industry developed own outcome. (S67)

Lack of ability to ensure that these outcomes will be used in clinical trials. (S75)

COS developers also saw the need for their COS recommendations to be updated, and seemed to be concerned that their recommendations had quickly become out of date:

Dynamic field where outcomes of interest change. (S15, published 2007)

Newer outcome measures and domains have developed since then. (S21, published 1997)

Outdated at this point. (S51, published 1999).

COS developers listed many process related limitations, including not addressing all intended issues (S6) and not providing a forum for external input to the consensus process (S35). Two COS developers listed '*no gold standard*' in relation to method for development as one of the main limitations (S2 and S46). COS developers also included some limitations regarding how to measure outcomes including that '*not all domains had instruments available*' (S52), and '*limited data about change sensitivity of instruments*' (S65).

5.5.7 Resources

Can you please list the resources you used to develop your core outcome set?

This was a free text response, and 77/81 respondents answered this question. Resources largely included funding (n=50), experts (n=11), and literature (n=4). The funding received varied between COS developers:

> *Funding for a researcher, funding for travel expenses, funding for international meeting, including flying people in from Australia and New Zealand. Was not massive, but I think around £30K for the overall thing. We were very lucky to have the dedicated support of a lot of people. (S61)*

> *WHO funding for one conference meeting - thus very limited resources. (S66)*

12/77 indicated that they did not have any resources. Four COS developers wrote that the only resource they had was their own time. Three other COS developers also noted time of experts as one of the resources used.

How long did your study take, from planning to completion?

Respondents were asked to enter the number of months, and 77/81 answered this question. 6/77 did not provide a number of months because they could not remember, did not feel it was applicable or because the work was still ongoing. The number of months ranged from 3 to 252; the median was 12 months, and the number of months taken are shown in Table 33**.** Studies that took the least amount of time tended to involve a single meeting, and solely considered what should be measured. This was in contrast to those studies that took the most amount of time that seemed to involve multiple methods and often considered how to measure the outcomes once they had been decided.

Table 33: Number of months to develop the core outcome set

Months	Frequency (n)
Less than 12 months	15
12 - 23 months	28
24-35 months	13
36 -47	9
48+	6

We asked COS developers to indicate whether the time taken was longer than expected, as expected, or shorter than expected. 54/77 recorded that this was as expected and one COS developer recorded this was shorter than expected; COS developers attributed this to efficient planning and having a deadline to meet, for example the length of a grant.

22/77 recorded that the time taken to develop the COS was longer than expected and described '*lack of time due to lack of funding*' (S26), difficulties with international participation, and difficulties with patient involvement:

> *Large group of individuals and authors to provide feedback, writing and edits over many timezones throughout the world...all working as volunteers on this project. (S62)*
>
> *We underestimated the difficulty of patient accrual. (S4)*

One COS developer described '*a lack of familiarity*' with the process (S23). Another COS developer simply wrote '*This stuff takes time*!' (S18) and another noted "*Difficult task, and probably not ever going to be done*" (S17).

5.5.8 What next?

Was the future implementation or uptake of the core outcome set considered by your group at any stage?

76/81 respondents answered this question, which had a yes/no response option. If respondents answered 'No', they were asked to explain why this was not a consideration, and if they answered 'Yes' they were asked what plans they had/have to promote the uptake of the core set.

Of the 12/76 (16%) who answered 'No' to this question, their responses indicated that they saw implementation as beyond the scope of COS development, or that funding was not available to support the implementation or uptake of the COS.

One COS developer described not having a way to look at the uptake of the COS:

> *We had no way of monitoring the uptake as these would be used in clinical trials most of which are proprietary. We will know over the next several years when the studies are completed. (S75)*

64/76 (84%) answered 'Yes' to this question. COS developers who answered yes mostly listed publication in a journal as addressing the issue of uptake (n=18), but some also listed participation in meetings (n=11), talking to relevant stakeholder groups (n=7), and interestingly, the involvement of prospective users in the development process who might influence uptake later on (n=5).

Three COS developers cited uptake in guidelines as a way of implementing the COS, this included FDA and NICE guidelines. Two COS developers also wrote about research funders' involvement in the implementation or uptake of COS:

> *Federal grants have required their use in recent years. (S10)*

> *The major sponsoring professional society distributes, references, and promotes the use of the work. (S43)*

Do you have any plans to update or review your core outcome set?

This question had a fixed yes/no response option and 76/81 respondents answered it. 36/76 (47%) answered 'No' to this question. These respondents indicated either that they had not thought about the need for updates or reviews of their COS or, as with implementation, indicated lack of funding and resources as a barrier to updates or commented that updates were not part of the task.

40/76 (53%) answered 'Yes' to this question. Reasons given for plans to review or update a COS varied and included:

> *To reduce size of COS. (S7)*

> *According to [OMERACT] Filter 2.0 all COS must have an update cycle. (S25)*

Review in a larger European group. (S65)

Some COS developers planned to update their recommendations to ensure relevance and '*to confirm importance of the domains to patients affected*' (*S35).*

COS developers also suggested when planned updates might take place, typically between 5 and 10 years, but they did not provide any explanation for their choice of time points.

5 year update cycle based on any new scientific publications, new evidence. (S31)

Originally planned to review in 3 years but due to poor results from recent clinical trials and number still underway it is best to wait until more data to review. (S45)

They will be revised approx. every 10 years. (S67)

Assess impact and revise in a 5 year time. (S74)

5.5.9 Guidelines

Reflecting on your experiences of developing a core outcome set, are there any areas that you feel would benefit from methodological guidance or research to inform future activity to develop core outcome sets?

Respondents were asked to list up to five areas. 78/81 respondents answered this question. Three of the 78 did not think this was applicable to them, and two of the 78 answered 'don't know.' Sixteen answered 'no' to this question. 92 areas for guidance or research to inform future activity were listed by the remaining 57 COS developers, but 17 of these were not relevant to COS development. A list of classifications can be found in Appendix 9, and a summary of the frequencies of categories are shown in Table 34.

Table 34: Areas that COS developers feel would benefit from methodological guidance or research to inform future activity to develop COS

Category	Frequency (n)
How to measure core outcomes	17
Stakeholder involvement	16
General	13
Choice of methods	6
Consensus methods	8
Implementation	4
Systematic review	3
Outcome terminology	2
Review and feedback	1
More data	1
Quality assessment	2
Wording (phrasing of outcomes)	1
Application of COS	1

COS developers made general comments about the need for guidance in COS development (n=13):

> *We've now developed OMERACT Filter 2.0 that is highly relevant to all COS developers. This, or something strongly like it, should be adopted by COMET ASAP. COS developers need concrete guidance NOW. (S25)*
>
> *Absolutely, any additional guidance possible offers best hope of developing appropriate guidelines (S31).*
>
> *I think guidance needs to reflect the practicalities of making decisions based on imperfect information. (S61)*
>
> *Guidelines for optimizing the process. (S67)*

COS developers felt that guidance or research to inform stakeholder involvement would be beneficial to inform future activity (n=16). Some made general comments about the need to involve multiple stakeholders and perspectives; others made specific comments about guidance:

> *Structured method for selection of experts and reaching consensus. (S27)*
>
> *How to identify participants. (S36)*

Patient involvement seemed to be a particularly important area where it was felt that guidance or research was needed. As well as the need for their involvement, COS developers suggested:

> *How to better incorporate patients at an appropriate level and time. (S62)*

> *More discussion about including minority patient participants in the process. (S44)*

One COS developer also commented on the need for *'guidance on wording of outcomes and domains when including patients in the process*' (S44).

One specific area suggested by COS developers as requiring guidance was choice of methodology (n=6):

> *Define different approaches and pros and cons of each. (S8)*

> *Which method to use (delphi, focus groups etc). (S36)*

Consensus methodology was another area where it was felt that guidance or research was needed (n=8). One COS developer asked '*what is consensus*?' (S7). Specific comments related to how to define consensus:

> *Noticed with others developing core sets - lots of confusion about consensus levels for Delphi/ NGT. (S44)*

COS developers also made general comments in relation to how to measure outcomes (n=17), although specific areas requiring guidance were not described.

5.6 Discussion

5.6.1 Main findings

We found that COS have been developed for a range of different reasons, but predominantly because of heterogeneity in what is being measured and heterogeneity in how outcomes are being measured. This thesis focussed on studies that aimed to establish what to measure, but as we also found in the systematic review of COS, consistency about how outcomes are measured is also important to COS developers. This highlights a key message, that COS development does not stop at what should be measured.

The absence of a COS was an obvious rationale for core set development, but interestingly the survey highlighted that it is possible that a COS existed but it was not deemed good enough or suitable for use. This highlights two important findings. First, although a quality assessment tool does not currently exist for COS development studies, users are making some informal quality assessment and applying some criteria as to what they think makes a COS of good quality, such as patient involvement, and the level of methodological detail provided. This affirms the need for an agreed method to quality assess COS studies. As the systematic review in Chapter 3 highlighted, and in view of the continuing increase in the number of COS being developed, this needs to be a priority in COS research.

The second area of COS development highlighted here is the importance of defining the scope of the COS. A COS may be developed rigorously but may not fit the required scope. It is therefore important that scope is well defined to allow users to decide whether a COS matches their required scope. Failure to do so could result in the COS not being used, the COS being used inappropriately or even duplication of work to develop a similar COS for a specific scope. This is therefore an important criterion that should be included in reporting standards and any reporting guidelines for these types of studies to allow users to make this kind of decision. COS developers' comments also touched on other questions about the scope of COS. As one COS developer wrote: 'Two core sets were available but they were both too broad.' Whether COS should be more general or quite specific is an interesting area for discussion, and may have implications for the uptake and implementation of COS.

In this survey, choice of methodology was most frequently attributed to the influence of previous work on COS development (by looking at the literature) or expert advice, partly confirming what was found in the systematic review of COS that COS developers have a propensity to use methods that have been used previously. This raises the question whether COS developers critically reviewed these methods when deciding which to use, and if not, this might not be the best way to make decisions if previous work is flawed or has not been methodologically rigorous. More detail about previous work used to inform decisions was derived from the survey, and included work by OMERACT and HOME. The COS developer of one recent study also cited the COMET website as a resource to help them make decisions about methodology. In contrast to what people are doing presently, the survey informed us about how COS developers have made decisions regarding methods in already published work. This highlights the need to find out from current COS developers

how they decide what methods to use; this has been explored during the qualitative interviews as will be described in Chapters 6 and 7.

The results of this survey show that COS developers regard patient involvement as an important area of COS development, but some also found this challenging. The involvement of different stakeholder groups is a vital part of COS development, as the survey showed that 42% of studies that examined outcome prioritisation by stakeholder groups observed between group differences. Of seven studies that had involved patients, three (43%) reported differences between patients' and clinical professionals' views. Half of the COS developers who selected 'experience of living with/caring for someone with a condition' and 'being able to see things from the patient perspective' as reasons for their choice of stakeholders did not include these groups in the process. These COS developers might have thought that clinical experts could provide that perspective without involving patients themselves. However, such a position is controversial and open to challenge. Certainty those COS developers in this survey who did involve patients described them as having different 'priorities' to other stakeholder groups, particularly clinicians, and placing 'more emphasis' on different outcomes. This exemplifies the importance of the choice of stakeholder groups to include in COS development, and the necessity of including patients in the process. These findings raise a further methodological question about how stakeholder group differences in outcome priorities are reconciled in the process of COS development, which was described as a challenge of the process by COS developers.

Intriguingly, a lack of data to support expert opinion was another commonly described challenge, as well as being the most frequently cited limitation of COS work. Little explanation surrounding this was provided by COS developers and due to the nature of the survey method, it was not possible to explore this issue further beyond their initial responses. COS development is essentially opinion-based, but interestingly COS developers seem to be implying that it could be something other than this. In the context of COS development, any data that is available will inevitably compromise different people's views regarding what is important to measure. Systematic reviews of outcomes used in published trials might be regarded as data or evidence to support decision making, but they only indicate what outcomes previous researchers have felt important to measure and therefore cannot be an 'opinion free' test or gold standard of what is an important outcome. It would seem that COS developers in this study expressed an unease about

opinion-based decision-making rather than it being a limitation of COS development per se. Evidence versus opinion is an important methodological consideration, and one that we need to explore further in this area of COS development.

COS developers reported being challenged by the lack of resources for their studies, which included limited funding, time and technology. One COS developer hinted at the implications of this challenge, commenting that '*speed meant we made a few mistakes*,' which suggests that resource issues have very real implications for the quality and outputs of the process. This was echoed in responses regarding the time taken to complete the COS development work, as close to a third of respondents indicated that it took longer than expected and a lack of funding was amongst the reasons given for this.

COS developers indicated that the uptake and implementation of the COS could be a strength, a limitation, and a challenge of the process. When asked directly about implementation as part of the process, however, COS developers predominantly listed publication in a journal as addressing the issue of uptake, despite this likely being only one step in ensuring that a COS is actually used. Nevertheless, implementation was described as challenging because it required changes to practice that would be hard to achieve. Respondents also answered that there was currently '*no way of monitoring the uptake*.' COS developers lacked the influence '*to ensure that these outcomes will be used in clinical trials.*' These are important obstacles to tackle if COS are to be successfully adopted in trials and research. Where COS developers had not considered implementation it seemed to be because funding was not available, or it was not felt to be part of the task. Overcoming these barriers is a priority for groups, such as the COMET Initiative, working in this area of COS development.

5.6.2 Robustness of the study

Strengths of the study

The study had several strengths. University of Liverpool software was utilised which allowed the survey to be designed specifically for the purpose of this study. This was important as the software allowed the inclusion of filter questions, which, in turn, facilitated the inclusion of specific questions to gain detailed responses when particular answers were selected. The online survey was developed in line with design principles

suggested to increase survey response and included both closed and open questions, which again allowed for specific and more elaborate responses.

The survey achieved a relatively high response rate (53%) for an online survey [110, 111], where rates of under 20% are not uncommon [112]. One possible explanation is that the survey involved a small specialised population [111]. As already noted, the steps taken when designing the survey to maximise the response rate could be another explanation for the high response rate achieved here. Furthermore, it achieved a high completion rate of 85% and the respondents were representative of the systematic review in relation to the disease categories, population and intervention characteristics covered.

Response rates were higher for those COS developers from the subset of more recently published studies, meaning that the survey results reflect more recent COS development practices than older. Although it would seem that COS developers were more likely to respond if they had involved patients than if they had not done so, this is likely related to response being higher for more recent studies as patient participation is more frequent in more recent COS. Given the survey's overall goal of investigating how methodological guidance or research can inform future activity to develop COS, this might be considered a strength of the survey. Recall is also likely to be more accurate for recent studies.

Limitations of the study

Thirteen COS developers started to respond to the survey but did not finish it, and the reasons for partial response are not known. There could be issues around the suitability of the questions, the length of the questionnaire or issues with respondents recall of the COS development process as some of the studies were published quite a time ago. These issues could also apply to non-responders, where again the reason for not responding is not known. Furthermore, although the results of the survey relate to studies that were published relatively recently, nevertheless 66/81 (82%) of the studies were published between 2000 and 2013, which is still some time ago. Recall bias may have affected the responses. Geographical location of the developers was not sought in this study, but could be considered as another factor that might be associated with whether a response was received or not.

A survey of this sort provides only a limited insight into participants' opinions and does not allow clarification or more information to be sought. The survey relates to work that has already been done, and was conceived and designed several years ago. COS development appears to be a fast-moving field, and it is possible that things might have changed or moved on since these studies were conducted. As these are not the most up to date COS studies, one might question the relevancy of these survey findings today. However, to date limited methodological guidance has been published in the area of COS development; hence it is expected that these findings will still be relevant for studies published more recently, and indeed for studies underway or in the future.

5.6.3 Summary

This survey is the first to provide insight into COS developers' methodological decision making, and choices of methods. It also provides insight into their experiences' of doing this type of work, including the challenges that they encountered and it gives the first account of COS development in a cohort of studies. The majority of respondents (73%) felt that there is a need for methodological guidance or research to inform future activity to develop COS, including: stakeholder involvement, patient involvement, choice of methodology, and consensus formation. These findings raise important methodological questions and highlight areas that future research and discussion should focus on. The main limitation is that the survey was carried out with COS developers who had already published this work, and as such, the results relate to work that was done several years ago. This highlights the need to find out from current COS developers how they decide what methods to use, and the challenges that they encounter. Interviews with recently published and ongoing COS developers, to explore these issues, are described in Chapters 6 and 7.

Chapter 6: Qualitative methods to study researchers' experiences of developing COS

6.1 Background

Investigating researchers' experiences of developing COS and the influences on the choices they make over the course of their projects is important. COS development is a new area of research and the formulation of guidance in this area depends on understanding what influences the methodological choices being made. As already described in Chapter 5, this thesis presents a mixed methods approach to explore COS development, drawing on an online web-based survey (chapter 5) and qualitative methods. The qualitative work will be the focus of this and the subsequent chapter.

6.1.1 Aims

I carried out qualitative interviews to explore how COS developers described their choice of methodological approach, including the influences on how they developed the COS. This aimed to identify potentially important but previously unanticipated issues as well as to describe in detail the COS developers' accounts of their projects. I also wanted to learn about COS developers' experiences and the challenges they encountered to identify priority areas for future methodological research.

6.1.2 Rationale for choosing qualitative methods

I drew on qualitative methods to examine researchers' accounts of their work to develop COS. Qualitative research is widely acknowledged as an appropriate method in evidence-based healthcare for providing in-depth insights into areas of practice, particularly when little previous work has been conducted in a particular area [113]. For a new area of research where little is known about the nature of the phenomenon, or few definitive hypotheses exist, qualitative inquiry is a reasonable starting position because it is exploratory. Qualitative research focuses not only on the meanings that people attach to experiences, but the relationship between experience, knowledge, social factors and the

actions taken that shape these processes. Understanding this is paramount to the aims of this study to further our methodological understanding of COS development processes, as it is through multiple and contrasting accounts that theory and understanding develop.

6.2 Reflexivity

Within the field of qualitative research it is necessary to be reflexive about one's role in that research and the influence of beliefs and behaviours on the research process. This is important because in qualitative work it is acknowledged that bias cannot be eliminated and that the researcher will inevitably influence the research findings. In this research I aimed to achieve an 'empathic neutrality', that is, to achieve a middle ground between becoming too involved which can cloud judgement, and remaining too distant which can, for example, stymie participants' accounts and thereby reduce understanding [114]. I therefore strove to avoid obvious, conscious or systematic bias and to remain as neutral as possible throughout the design of this study, as well as the collection, interpretation and presentation of data [115]. However, it is important to recognise that research will be influenced by the researcher, so in turn, it is important to reflect on my own personal characteristics and how these could have influenced the current study.

I recognise that my age, gender, training and qualifications will differ from many of the interviewees (COS developers) and may therefore have implications for the way in which they perceived me. I have undertaken training in qualitative research, specifically around conducting and analysing qualitative interviews (see Appendix 10 for a full list of relevant training). I have a background in psychology and obtained a First class BSc in Psychology and Health Science from The University of Liverpool in 2007. After graduating, I went on to work for a patient reported outcomes (PRO) consultancy specialising in the development and use of PRO endpoints in clinical trials, before joining the University of Liverpool in 2009 to work on a systematic review of paediatric adverse drug reactions.

It seemed particularly important to consider how my current professional role might have influenced this study. I have been involved in the COMET Initiative, as the Project Coordinator, since its inception in 2010. As part of my role I have been involved in multiple COS publications, organised and participated in COMET conferences, and raised awareness about the COMET Initiative all around the world. I have been a point of contact for COMET, including when COS developers have registered work in the COMET database. These roles

could have influenced the way in which interviewees engaged with me and how I gathered and analysed data. For example, it is possible that interviewees perceived me as an authoritative figure in the area of COS research because of my association with the COMET Initiative, and as such could have been reluctant to share poor practices with me. In order to manage the potential impact of these issues on this work, I tried to position myself very much as the student in the interview, and the interviewee as the expert with experience in the field of COS development that I wanted to learn from. To facilitate this positioning, I included the following wording in the invite email: 'You have been asked to take part in this study because you are or have been involved in research to determine the outcomes or domains to measure in clinical trials, and your experiences are therefore very important to us. We would like to learn from your experiences and incorporate what we learn into guidance to assist future researchers.' I also repeated this at the start of each interview.

It is also worth noting that my status as a PhD student could have been influential and interviewees could have withheld information that they felt I might not understand. Coming to this as a PhD student, it would be possible to feel overwhelmed or intimidated by interviewing some of the more senior figures who have led COS development in a number of areas. To avoid that being an issue, I tried to establish rapport with interviewees prior to the interview itself, through the enrolment process, thereby putting myself at ease prior to conducting the interviews. Furthermore, during my time as the COMET Co-ordinator, I had gained experience of corresponding with senior figures so for me personally, interviewing people in senior positions was not something that I found concerning or intimidating.

6.3 Methods

As no previous qualitative research had been conducted on COS developers' accounts of their work, in this study I adopted an interpretive approach informed by the constant comparative method [116]. An interpretive approach places emphasis on understanding rather than simply an explanation of the phenomena under investigation, which is imperative to the aims of the work undertaken in this thesis. In the analysis I therefore went beyond simply describing what participants said, to interpret their accounts and consider what could be learnt from how they constructed their accounts. The constant comparison method involved taking one piece of data (e.g. one interview or one theme) and comparing it with others that may be similar or different to conceptualise the data and

explore relations between various pieces of data [117]. These methods were particularly suited to topics that have not been previously explored because they help to avoid imposing assumptions, and allow a close connection between the data and its conceptualisation [107].

6.3.1 Interviews

I conducted semi structured interviews, a fairly open interview format that allows for focussed yet conversational communication. I felt that such interviews allowed specific issues to be addressed whilst at the same time this approach allowed interviewees to freely describe their experiences, thereby eliciting information that might not arise in a structured interview and giving insight about what COS developers deemed to be important. A semi structured interview also allows flexibility to ask new questions that follow up interviewee's replies, as well as changing the order of questions to best suit each individual interview [107].

While the interviews were therefore conversational, I used a topic guide comprising key issues and subtopics to be explored with interviewees. The topic guide was used as an aide memoir to ensure consistency in data collection. Consistency here does not mean asking the same questions in the same way to each interviewee, but rather steering the general topics for data collection in each interview. The topic guide was informed by the results of the systematic review. For each interview there was a list of questions and fairly specific topics to be covered (see version 1.0 of the interview guide in Appendix 11). To avoid generalised and idealised responses which might have arisen if questions focussed only on opinions, the topic guide focused initially on COS developers' experiences of the processes involved in developing a COS. Open ended questions were designed to elicit COS developers' accounts of what they did to develop a COS, what influenced their decisions about their approach, and what constraints they were operating under. I reviewed interviewees' relevant publications prior to each interview in order to tailor questions and prompts to each individual being interviewed. I also avoided phrasing questions in ways that might have made interviewees feel their judgement was being questioned or challenged. For example, if interviewees did not talk about including patients as participants, I used an open question such as 'Did you think about including patients in this work?' as opposed to 'Why didn't you include patients?' which might seem challenging to the interviewee.

By and large, all topics in the guide were explored with every interviewee. I asked additional questions, not included in the guide, to follow up issues as these were identified. The topic guide was also developed iteratively in light of the data analysis, which was conducted in parallel with the interviews. This iterative process allowed respondent validation whereby I was able to check the correspondence between the developing analysis and the perspectives and experiences of interviewees in subsequent interviews. Furthermore, iterative analysis allowed identification of unexpected or atypical issues, and provided the opportunity to explore these issues with subsequent interviewees. An example of an unexpected issue that arose during the interviews was around ethics requirements for COS work. This was not included in the original topic guide, but once it was identified, I was able to incorporate it into the guide to explore the issue in the interviews that followed.

Interviews were conducted via telephone. Due to varied geographical locations of COS developers, face-to-face interviews were not feasible for this study. The benefits of face-to-face interviews are well documented. However, there is no evidence that the nature, depth or quality of response is diminished in telephone interviews, when compared to face-to-face [118]. Furthermore, telephone interviews may actually minimise socially desirable responses, and reduce bias or distortion due to interviewer's characteristics [119].

All interviews were conducted in English. As some participants were overseas, I ascertained the feasibility of doing an interview in English with the participant beforehand. Decisions were made on an individual basis, and if there was doubt that the interview could be conducted in English then the interviewee was asked to nominate another potential study author to interview. I asked interviewees for permission to go back to them after the initial interview to follow up areas that had not been covered in the interview or if new questions arose as analysis developed. I made field notes after each interview as an aide-memoir in which I recorded contextual details, my personal reflections about the interview and my initial analytic thoughts [107].

All interviews were audio recorded, transcribed and anonymised. I carried out a small amount of transcription to inform the transcription protocol. The majority of the recordings were then transcribed by a professional transcriber, who was asked to sign a declaration of

confidentiality and instructed to adhere to the study protocol for data handling and storage.

6.3.2 Participant selection and recruitment

Purposive theoretical sampling, a form of non-probabilistic sampling, was conducted to discover categories relevant to the research questions [107]. Sampling in qualitative research is non-probabilistic because, with relatively small sample sizes, probabilistic samples would likely be too homogeneous to be of much use. Purposive sampling refers to a process where participants are selected because they meet criteria that have been anticipated by the researcher as relevant to addressing the research question.

I used quota sampling to achieve purposive sampling. Quota sampling involves selecting a sample that reflects different criteria or characteristics of the population in question [107]. Unlike stratified sampling, it is not done randomly, rather the interviewer selects people who fit the chosen criteria, in this case, I sampled to reflect the different aspects of COS development identified through the systematic review of COS.

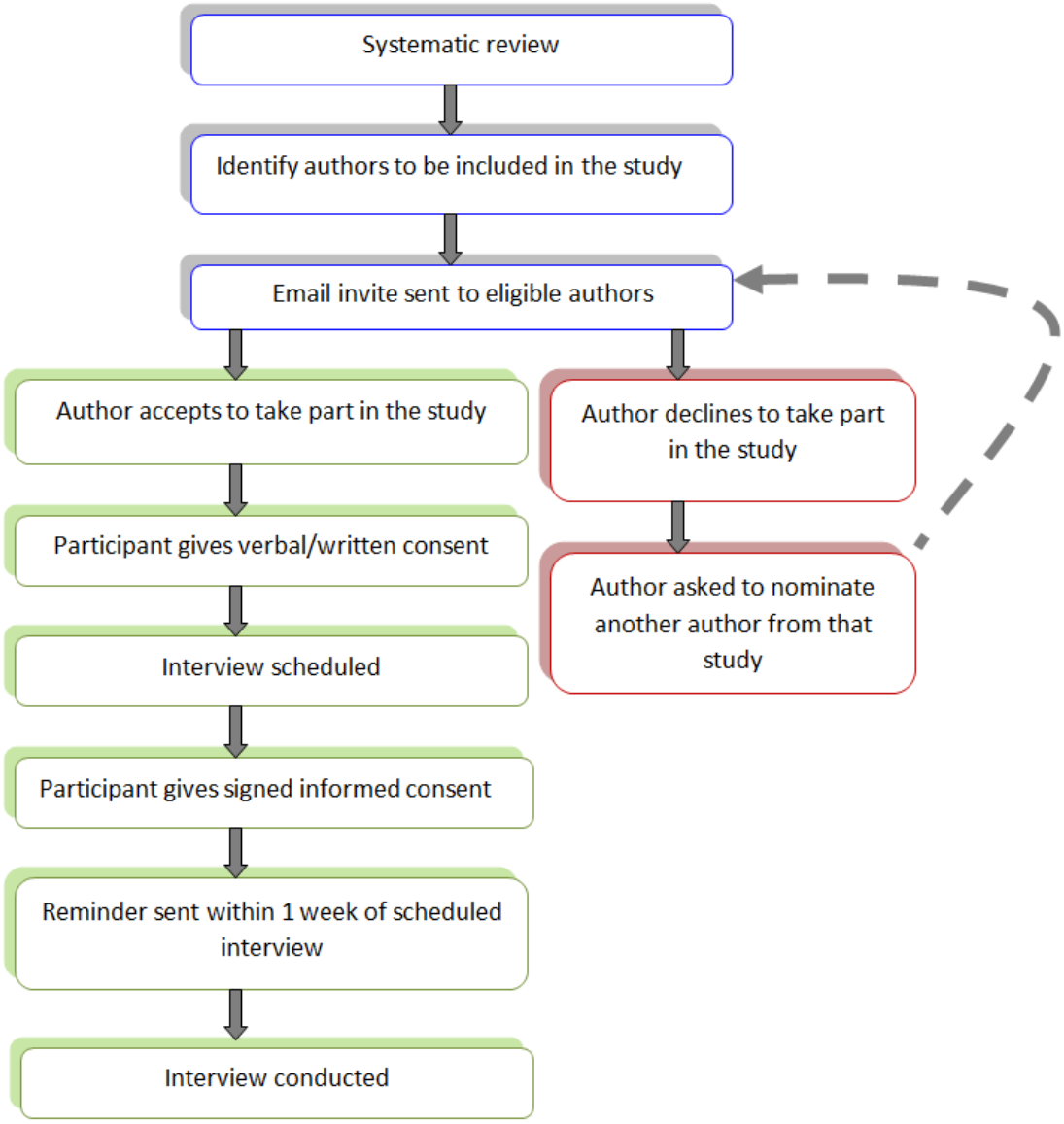
The sampling was also iterative. This type of sampling is an ongoing process whereby the researcher collects and analyses data in parallel, and uses the developing analysis to inform the sampling with the goal of enhancing the development of themes and theory.

The participant selection and recruitment process is summarised in Figure 7. I aimed to interview the lead author of selected studies. When this was not possible, secondary authors were considered for inclusion based on their level of involvement and responsibility for the work and publication. The final decision as to which author was interviewed was decided on a case-by-case basis.

Sampling aimed for theoretical saturation, so I continued conducting interviews until no new insights were being gleaned from the data. The very nature of theoretical saturation means that the sample size could not be defined in advance, but given the relatively large number of categories in my sampling matrices, I anticipated that I would need to conduct between 30 and 50 interviews to achieve saturation. As 30 is considered a ‘medium’ size sample for a qualitative study [120], I considered this to be a realistic number of interviews to conduct in the time available.

As already noted, I aimed to conduct interviews with the lead author. I contacted eligible authors by email to invite them for interview (see Appendix 12) and seek consent (see section 6.5.2 Informed consent). If the author declined, I asked them to nominate another author from the same study who was then invited by email. If an author did not respond, two further emails were sent out, two and four weeks after the initial email. I recorded authors who had not responded six weeks after the date of initial emails as non-responders. I contacted authors who agreed to be interviewed, by email, in order to confirm their acceptance to participate, and to schedule a convenient time for the interview to take place. I also confirmed that I had all relevant publications prior to interview and sent a reminder one week before the scheduled interview date.

Figure 7: Participant selection and recruitment process



I devised the quota sampling matrices for this study, shown in Table 35 and Table 36, to inform sampling of COS. The systematic review of COS, described in Chapter 3, enabled me to identify what I refer to as aspects of the COS development process. These aspects provided the starting point for informing the sampling of COS developers who were the interviewees for the qualitative study. The aspects represent features of COS development that I thought were important to explore in detail; how I sampled for them and why I thought they there were important are summarised below. Although the systematic review provided a starting point for the work described in this chapter, it did not constrain the qualitative work that was undertaken and described herein. Furthermore, compared to the survey, I orientated the qualitative study to focus on COS development studies that were ongoing or had been published more recently, in order to understand current COS development. Within these categories I used random sampling to select the particular COS to be included in this study. The goal was not to reach saturation across all of these groups, rather to explore characteristics within groups.

There were 86 unpublished studies listed in the COMET database* (date searched 20/01/2015), of which 57 were listed as 'COS' studies. Fifteen of these had a planned completion date of 2014, so I targeted these as the most likely to be nearing completion and therefore suitable to provide insight into the current approaches to COS development. Studies nearing completion were particularly targeted because the COS developers would have experience of different stages of COS development. I also enquired with the members of the COMET Management Group to obtain details of any other studies they knew to be near completion so that we could identify other potential COS developers to be interviewed. Due to the small number of ongoing COS nearing completion in 2014 to choose from, I invited all developers in this category to be interviewed.

** The COMET database is described in more detail in Chapter 2, section 2.1.*

Table 35: Sampling matrix

Aspects of the COS development process	**Methods** **n= number of COS**								
	Semi structured group discussion			**Consensus development conference**	**Nominal group technique**	**Unstructured group discussion†**	**Delphi**	**Lit review/ SR**	**No methods**
	Workshop	Meeting	Round table						
People involved									
Clinical expert (no PPI)	2-3	2-3	2-3	2-3	2-3**	2-3	2-3	2-3	2-3
PPI	7-12*^					1***	2-3		
Aim of study									
Wider trial design issues	15-25								
Outcomes only	15-25								
How/what to measure									
What to measure only	9-12								
What to measure + discussion/consideration of how but no recommendation	9-12								
What + how (done together)	9-12								
What + how (done in two stages)	3-5**								

Mixed methods

6-9 COS using single method, 6-9 COS using mixed methods

6-9 COS using single method, 6-9 COS using mixed methods

† Descriptions included task force, work group, working group/party, committee, board, and panel

^ I have grouped studies that included a face-to-face meeting (of any type) due to small numbers of studies that included patients, which made it hard to sample by each individual method

** Only 7 studies in this category were 2010 or later, so I also included some pre 2010 studies (targeted most recent)*

*** 1 or less in this category were 2010 or later, so I included pre 2010 studies (targeted most recent)*

**** Only 1 study in this category in the review*

Table 36: Other COS aspects sampled

Year of publication	Number of COS N
Pre 2010	3-5
2010-2013	17-25
Ongoing	10-15
Scope	
Population characteristics – Age	
All	3-5
Children	3-5
Adults	3-5
Not specified	15-30
Intervention	
All	3-5*
Drug	3-5
surgery	3-5
Specific (other)	3-5
Not specified	12-25
Number of COS/disease category	
Clinical area where more than 5	15-25
Clinical area where less than 5	10-15
Funding	
Commercial	3-5
Non-commercial	3-5
Commercial and non-commercial	3-5
Not reported	15-30

Disease categories

At least 1 COS from each category (10 categories with more than 5 COS)

At least 1COS from 5 different categories (10 categories with less than 5 COS after 2010)

** 1 or less COS in this category were 2010 or later, so I included pre 2010 studies (targeted most recent)*

6.3.4 Aspects of the COS development process that informed sampling

Methods used

The results of the systematic review demonstrated that a variety of methods have been used to develop COS, and that it is currently uncertain which of these are the most suitable, feasible and efficient. The results suggested that the methods used in the development process may influence the conclusions derived.

For example, consensus work was unknowingly undertaken by three different COS development groups in the same clinical area (paediatric asthma), but each group

identified different outcomes as core, suggesting that the use of different methodological approaches and stakeholders by COS developers may influence the outcomes included. Reddel et al [104] reported work undertaken by The American Thoracic Society (ATS) and European Respiratory Society (ERS) involving round-table discussions comprising clinical researchers, pharmaceutical representatives and regulatory agency representatives, with the aim of recommending outcomes to include in clinical trials of therapies for asthma in adolescents and adults. Busse et al [105] reported a workshop convened by National Institutes of Health (NIH) and the Agency for Healthcare Research and Quality comprising representatives from adult and paediatric asthma, pulmonology, and allergy/immunology; as well as lay voluntary organisations, biostatisticians, guideline developers and health policy representatives and pharmaceutical representatives in order to propose which asthma outcomes should be assessed in future asthma clinical research studies. Finally, Sinha et al [103] carried out a survey with parents and young people with asthma and a Delphi survey with paediatricians and specialist nurses, to develop a method by which to identify outcomes of particular relevance when evaluating the effects of regular therapies for chronic childhood asthma. Although there were some similarities in the outcomes selected as core by each of the three COS development groups (two common outcomes), there was variability in the other core outcomes identified. Research to investigate COS developers' choices will help to illuminate the reasons for these differences in methodological approach, and how these lead to differences in the outcomes deemed to be important. In sampling COS developers to invite to take part in this study, I therefore sought to maximise the diversity of the different methods that developers had used in their work. I anticipated that maximising the diversity of the methods encompassed in my sample would allow me to identify and compare the reasons for choice of approach and, in turn, identify patterns in their rationales that could inform guidance development.

Stakeholders involved

The participants who are regarded as key to the development of all COS will likely vary between clinical areas, but two stakeholder groups that are always likely to be important are clinical experts and patients. The systematic review demonstrated that where developers had described the participating stakeholder groups (n=172 studies), almost all had included clinical experts (171/172 studies) but only 18% (31/172) included patients in the process. Historically, the outcomes reported for trials have not always reflected the endpoints most meaningful to patients. Examples exist where patients have identified an

outcome important to them as a group that might not have been considered important by practitioners on their own [103, 121]. Although it is increasingly recognised that including patients in the development of COS is important, their involvement has been limited to date. I therefore aimed to recruit a proportion of COS developers who had included patient participants, as well as some who had not, to explore the influences behind COS developers' choices about which stakeholders to include as participants in the development process, and to identify the barriers and facilitators to the inclusion of different groups. As few studies in the systematic review of COS included patient participants, I also wanted to find out about published COS developers' experiences with patient participation as in some senses they could be considered as 'innovators'; it is important to understand the reasons for their efforts to seek patients' input and the challenges this brought, if guidelines are to facilitate patient participation in future COS development.

COS study aims

In the systematic review of COS, half the studies considered outcomes whilst addressing wider clinical trial design issues (51%), while half specifically considered outcome selection and measurement (49%). I included both study types in my sample of interviewees in order to explore the ways in which this consideration might influence the process of COS development and the methods used.

In sampling interviewees I also took into account whether the COS developers had focussed only on 'what' outcomes to recommend for a core set, or whether they had also looked at 'how' to measure outcomes, as I thought this likely to contribute to methodological decision making if multiple goals need to be achieved. In the systematic review (198 COS identified), 62% of studies made recommendations about what to measure only. Some of the remaining studies also made recommendations about how to measure outcomes, with 35% of studies doing this as a single process, i.e. considering both what to measure and how to measure in an integrated way. The remaining 3% of studies in the systematic review of COS considered what to measure and then how to measure outcomes as a two stage demarcated process. Although the focus of the qualitative work described in this and the subsequent chapter was on the process COS developers used to identify what to measure, I attempted to include COS developers who had used each of these approaches to explore any potential differences in methodological approach.

Scope of COS

Disease area

It is possible that how much COS work has been done in a particular disease area may affect how other COS developers in that particular area subsequently develop COS. The prevalence of COS work in a particular disease area might be useful in understanding methodological choices. I therefore aimed to include a range of disease categories with varying levels of COS activity to explore whether the volume of work in a given disease area influenced COS developers' choices about methods. Furthermore, by sampling a range of disease categories, I hoped to explore issues that might apply to particular patient groups, and how the type of patient group (e.g. patients with dementia, brain injury, learning difficulties etc.) might influence COS developers' choice of methods.

Age of population

Specific populations might pose distinctive methodological challenges. The age of the population is one aspect that might bring unique challenges. For example, the design and conduct of clinical trials in children presents different challenges when compared to trials in adults [122]. These include difficulties in the selection, measurement and reporting of outcomes. The outcomes that are important to measure and report for adults may not be the same for children, and there may be outcomes that are unique to adult or paediatric populations. Furthermore, outcome measurement instruments developed and validated in adults will often not be appropriate for children or babies which presents methodological challenges when designing research with children [123]. Examples exist where careful consideration has been paid to whether certain outcomes should be the same or different for adult and paediatric populations with the same condition. One such example is the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) group, where developers carried out separate consensus work for adults and children, and provided distinct recommendations for each population [124, 125]. For example, the paediatric core set featured sleep as a core outcome but this did not feature in the adult core set. This demonstrates the need for clear consideration of the potential differences in outcomes between children and adults.

Within the systematic review the majority of studies (149/198; 75%) did not specify whether their recommendations were specific to children, adults or both. Thirteen (7%) studies explicitly stated that their recommendations applied to both children and adults. whereas other studies made recommendations explicitly for one population or the other.

Twenty-three studies (12%) specified that the recommendations related to children only, and 13 (7%) to adults only, with three of the latter specific to 'older adults'. Whilst there might be important or distinctive methodological considerations across the whole age spectrum, there were too few studies in the review to make older adults a specific focus in this study. I therefore wanted to explore whether there were different methodological considerations when developing a paediatric COS, and if so, what impact this had on the way the COS was developed. An example of a specific methodological consideration might be the potential challenges of involving children in the process of deciding which outcomes are important.

Intervention characteristics

The majority (115/198; 58%) of COS identified in the systematic review did not specify whether the COS was developed for all interventions or a specific intervention. Of those studies that did specify the type of intervention (83/198; 42%), only seven (8%) made recommendations that were clearly intended for all interventions. The remaining studies specified the intervention type to which the COS should apply, which included drugs (n=40), surgery (n=13) and vaccines (n=2). A recent article summarised the key problems with surgical trial outcomes as follows: choosing the right outcomes for the trial (considering design and purpose); selecting relevant outcomes to measure from a range of possible outcomes, and selecting outcomes with a minimal risk of bias [126]. These issues are similar to those faced by all trials; however, it is unknown whether the process of deciding which outcomes are important to measure and report are also similar across all trials regardless of intervention type. I therefore attempted to include COS developers who had not specified the intervention type as well as those who had, to try to understand whether and how this might influence the methodological choices made.

Year of publication

I oversampled studies published since 2010 as this was the year the COMET Initiative was launched, and to reduce potential problems with participants being unable to recall details of the COS development process, which might be a difficulty where COS had been developed many years earlier. Developers of COS where the study was published prior to 2010 were considered for inclusion if studies published after 2010 were not available within a particular sampling category. While it is important to understand how COS have been developed in the past, I decided to focus this qualitative study on more recent and

ongoing studies (i.e. unpublished studies) as my study aimed to inform future guidelines for developing COS. To do that, I need to understand the current environment of COS development.

Funding

Funding source is a known potential bias in clinical research, including trial design and data analysis [127]. For this reason, I wanted to explore whether the source of funding influenced the methods used to develop COS. 47% of studies in the systematic review did not report the source of funding. Of those that did, I categorised 21% as commercially funded which includes studies funded by pharmaceutical companies and other for-profit health-related organisations; 57% were categorised as non-commercial which includes COS funded by universities and affiliated organisations, governmental agencies, professional medical societies and organisations, and charitable organisations; while 22% were categorised as receiving both commercial and non-commercial funding. I wanted to include studies from each category, including where the source of funding had not been reported, to explore any influence that funding source might have had on the development process.

Plans to review/update COS

Finally, the majority of studies identified in the review (167/198; 84%) did not report plans to update their recommendations, while 31 (16%) referred to a plan to update their COS recommendations. It is plausible that COS developers who had plans to continue COS work might differ in their methodological approach. However, due to my decision to focus the qualitative work on more recent and ongoing studies, I did not think it was appropriate to sample for variability in terms of future plans to update COS work, as it was likely that insufficient time would have elapsed for developers to be thinking about updating their COS work.

6.3.3 Data analysis

Undertaking an interpretive approach to analysis necessitates iterative data analysis, so I began analysis soon after initiation of data collection.

I have attempted to provide an audit trail to illustrate how the data and the conclusions I have drawn about the data are linked [117, 128, 129]. As described in more detail below, I

used framework analysis to structure my approach to data analysis and management, and to facilitate the linkage between data and conclusions.

Framework analysis

I used framework analysis as an analytic tool to structure how I worked with the data and to support data management [115]. This technique includes indexing and sorting (described in steps 1 to 4 in Figure 8 below) that is common across many different qualitative approaches; this involves taking a set of descriptive themes, as well as subthemes identified through the data, and developing a framework that can be used for indexing the data. Framework analysis adds one further step to other qualitative approaches, namely data summary and display, whereby the framework forms a basis for constructing thematic matrices for presenting the data in outline. In the matrices every participant is allocated one row, and each column represents a separate subtheme. Data are then summarised by participant and by subtheme and the summary entered into the appropriate corresponding cell.

Figure 8: Five key steps of framework analysis

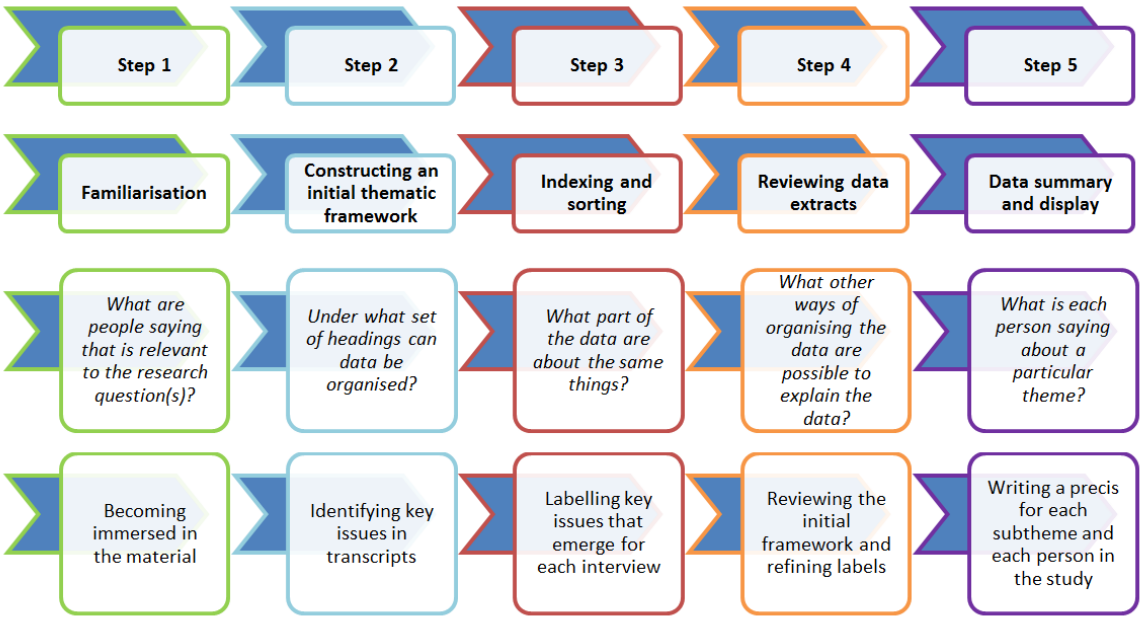


Figure 8 was developed based on the description of framework analysis by Ritchie and Lewis [115]

I selected framework analysis because it facilitated the constant comparative technique through the review of data across the matrix [130]. Furthermore, framework analysis allows both predetermined and novel ideas to come from the data. This was particularly useful as interview guide topics could be explored and developed without stunting the development of new themes or subthemes. The matrix format also facilitated sharing of data across a multi-disciplinary team, allowing individual team members to engage with summarised data and offer their perspectives without needing to read all transcripts; this was particularly helpful with the management of a large dataset. Finally, the framework approach helps to maintain links between the original data and findings. In turn, this facilitates transparency, adding to the rigour of the research process, and potentially, the validity of conclusions drawn from the data [131].

In the first stage of framework analysis I read the transcripts in order to gain an overall impression of the data. My supervisors also read the first six transcripts to help inform the analysis. I checked field notes for additional contributing information and to ensure the quality of the analysis. The next stage involved line by line analysis and the development of open codes to summarise the content of the transcripts. I then grouped the data together through a process of identifying recurring patterns and themes, and organising these into categories. I identified variations in the data and paid particular attention to outlier or 'deviant' cases, as these can help to 'test' and thereby develop the analysis. Coding was supported by a qualitative computer software package (NVivo version 10 [132]). I subsequently developed subtheme framework matrices for each theme identified. An example of an excerpt from a framework matrix is provided in Table 37. I referred to field notes when gaps were present at a participant level in these matrices, as a first check of whether the theme in question had featured in a particular participant's accounts, as well as to put data into the context of the full interview when clarification of their full account was helpful for interpretation. Furthermore, some themes or categories may only be identified by considering a participant's account across the whole transcript, for example what they do not say relative to other participants. Finally I summarised the themes by comparing responses in terms of similarities and differences, and these were then reviewed by, and discussed with, my supervisors (BY and PW). I used excerpts from interviews to support the summaries. In terms of notation for displaying excerpts, I use […] to denote text that seemed unnecessary in interpreting excerpts and which I removed in the interest of brevity.

Table 37: An excerpt from the 'Choice of methods' framework matrix

Identifier	The literature	Advice from others	Experience	Resources
P1	Decided to use Delphi based on the literature. 'and it's very frequently used in the, for developing core set of outcomes.' 'Not many methods to do it' - did consider doing a 'consensus method' meeting but decided 'it was not very systematic.'			
P2		Methodology people' (also calls them 'methodology team') advised them to use GRADE methodology - 'If you really want to in a (.) in a erm (.) ((sigh)) scientifically robust way, you know, or er er a process that you can argue is, is, somehow more than just your professional opinion, use this GRADE working group erm methodology. And so that was their recommendation and they walked us through it.' Influence on how the Delphi was run.		
P3			They took approaches that they were 'somewhat familiar with.'	
P4	I think very much influenced from the papers we had read about from OMERACT and they sometimes used this approach, so. ' Work of OMERACT influenced what they did - chose this approach because they had used it. So influenced choice of method.	OMERACT person 'told' them they needed to set consensus rules before voting and so on. So influenced how they conducted the Delphi.		
P5	Looked at similar work that had been done, especially in Rheumatology 'but there were others as well.' 'If we use the Delphi technique we can you know, systematically solicit info…' Shows they looked at what others had done but did think about it. Other people had used this approach and been successful, and the rationale for it was 'very convincing.' Used rheumatology work as a guideline and then asked 'can we replicate that work in our area?' But if more explicit guidelines existed they would have looked at them - suggests there was still uncertainty. Read books on consensus and papers.	Sought advice from others first who were not very helpful, then found a rheumatologist who shared their experience. 'He was a mentor for me. Um I met weekly with this guy for years.' HE would suggest work that had been done for her to look at. 'Well um, well he was my mentor and so every step we took he advised.' 'We talked through, well how might we do this, you know? Um what should I go read to learn more about this Delphi thing I never heard of? Um (.) you know, who could provide some advice about how this might, you know, how this could play out for us? All of that.' Talks about this as guidance as opposed to being told what to do.		Oh, well (.) number one we didn't have very much money, (.) so even though maybe it would have been lovely to have had some sort of consensus conference or whatever, there was no way we were going to get this um degree of expertise all in one place.' However, went on to say that this would not have actually suited their patient population. But it 'wasn't financially possible' to have done this for clinicians - 'if I'd had unlimited time and resources maybe that, we could have done that.'

6.4 Quality in qualitative research

This study was informed by the literature on quality in qualitative research, to ensure appropriate procedures were followed and that analysis remained grounded in the data. However, it is important to note that although these quality procedures can strengthen the rigour and transparency of qualitative research, they do not guarantee quality [133]. Furthermore, quality procedures require the exercise of judgement on the part of the researcher. Quality in qualitative research is inseparable from the researchers themselves, hence it is important to document the process of reflection [129, 134, 135]. My personal reflections have been discussed in section 6.2 of this chapter.

Quality can also be judged by what a research study contributes to an area and whether it offers insights that have the potential to change practice or research [136]. Sometimes termed 'catalytic validity', this refers to the usefulness of research and its potential for real world impact [137]. Catalytic validity is an important validity criterion given the aims of this study to identify priority areas for future methodological research and guidance in COS development.

I also considered the literature on standards for reporting qualitative research. Reporting on research design and methods of data collection and analysis highlight distinctive features of the research such as approach and researcher characteristics, but it is also important to justify choices to ensure that assumptions and decisions are transparent to the reader [138]. THE EQUATOR network recommend two guidelines for reporting qualitative research, the 21-item Standard for Reporting Qualitative Research (SRQR) [138] and the 32-item Consolidated Criteria for Reporting Qualitative Studies (COREQ) [139]. The guidelines are largely overlapping and therefore both were considered in the reporting of this study.

6.5 Governance, ethics and confidentiality

6.5.1 Ethics

This research study fully complied with the ethical practice guidelines laid out by the British Psychology Association. NHS ethics approval was not necessary for interviewing researchers who were also NHS staff. However, because it was possible that this study might also have included patients, if any had been found to lead COS work, I needed to

consider whether NHS ethics approval was required. After discussion with the Chair of the Institute of Psychology, Health and Society Ethics Committee, University of Liverpool, it was decided that the project did not require NHS ethics committee approval because I identified and recruited participants through published research and not through NHS facilities within which they may be patients or clinicians. Instead, approval was sought and granted through the University of Liverpool's ethics procedures (Research Ethics Subcommittee for Non-Invasive Procedures reference number: RETH000624).

6.5.2 Informed consent

All interviewees gave signed, informed consent prior to proceeding with the interview, after first being provided with an information sheet about the study (see Appendix 13) and a consent form (see Appendix 14). I explained how the study aimed to assist the research community by informing the development of guidance standards for COS development (see invite email in Appendix 12). Consent was sought remotely, via email rather than face to face, due to distance between research location and participants. The recruitment process allowed interviewees time to discuss the study with myself or one of my supervisors (PW or BY), as well as have any questions answered. Participants were informed that they could withdraw their consent and leave the study up until the point of interview transcript anonymisation, without having to explain or provide reasoning. Consent for audio-recording the interview was sought as part of the consent process prior to commencing the interview.

6.5.3 Data protection and anonymity

This research study complied with the Data Protection Act of 1998 concerning the handling, processing and storage of data. All data (consent forms, audio recording devices, field notes and interview transcripts) were kept in a secure locked filing cabinet or electronically in a secure filestore. Files were deleted from recording devices as soon as possible, and held securely (encrypted) until transcribed and transcriptions checked. On completion of the publication of the study, audio recordings will be destroyed. In accordance with the University of Liverpool Research Data Management Policy, all other research data (consent forms, anonymised interview transcripts, field notes, and contact details) will be kept in locked filing cabinets and/or password protected university computers for ten years.

All data included in the analysis and write-up of this study has been pseudo-anonymised, with the removal of identifiable features, such as place and person names. Details necessary for interpretation (e.g. clinical area) were retained in the transcripts as they could be important in explaining meaning but have been removed in the reporting of results (Chapter 7). The data from the interviews was transcribed as soon as possible following the completion of the interview. I assigned each transcript a unique identifier code. Published COS developer transcripts were allocated a code starting with P, followed by a number, and ongoing COS developer transcripts were allocated a code starting with an O. Myself and one of my supervisors (BY) had access to the pre-anonymised data (audio-recordings) as necessary for BY to advise on the quality of the interview process.

The results are described and discussed in Chapter 7.

Chapter 7: Qualitative study findings on researchers' experiences of developing COS

7.1 Introduction

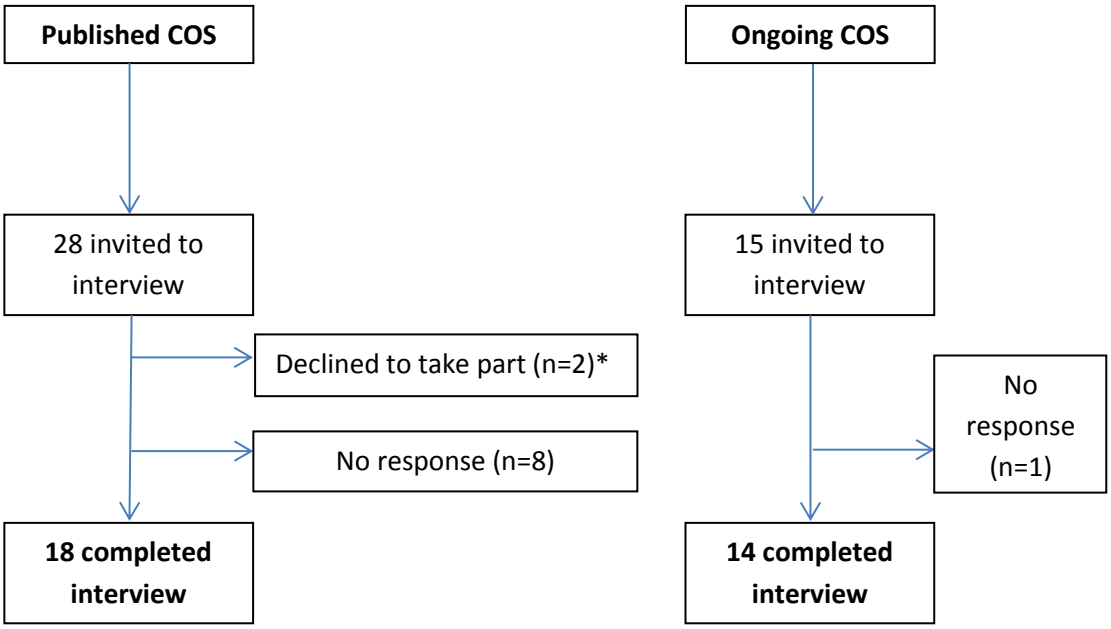
As described in Chapter 6, I carried out qualitative interviews to explore how COS developers described their choice of methodological approach, including the influences on how they developed the COS. This aimed to identify potentially important but previously unanticipated issues as well as to describe in detail the COS developers' accounts of their projects. I also wanted to learn about COS developers' experiences and the challenges they encountered to identify priority areas for future methodological research. The results of these interviews will be the focus of this chapter.

7.2 Results

I interviewed 32 COS developers (18 with published, and 14 with ongoing, COS projects) between May 2014 and June 2015. The study flowchart, showing inclusion of participants, is presented in Figure 9. Published COS developer transcripts were allocated a code starting with P, followed by a number, and ongoing COS developer transcripts were allocated a code starting with an O.

Interviews lasted between 33 and 120 minutes (median length 63 minutes). Published COS developers comprised 16 lead authors, one last (corresponding) author and one second author. Six were from Europe (33%), 10 from North America (56%) and two from Australasia (11%). Two of the published COS developers had worked on a COS previously. Ongoing COS developers comprised 10 principal investigators, three clinical leads, and one PhD supervisor. Of the ongoing developers, 12 were from Europe (86%) and two from North America (14%). Summary sample characteristics for the 32 COS are provided in Appendix 15 (N.B. as discussed in the methods in Chapter 6, the goal was not to reach saturation across all of these groups, rather to explore characteristics within groups)

Figure 9: Flowchart showing inclusion of participants



**One participant did not give a reason for declining, the other participant was 'over committed'.*

Participants spoke of how COS development was a particularly challenging area of research because it is an emerging field, without shared assumptions or an advanced knowledge base. As such, this chapter will focus on influences on developers' choice of methods, the challenges encountered or perceived throughout the process of COS development, and the areas of COS development where developers emphasised an absence of guidance.

7.2.1 Influences on methods

COS developers discussed a variety of influences on their choice of methods for COS development. Most discussed more than one reason for the methods they used, as summarised by O32 who said '*the Delphi was chosen for a bunch of different reasons.*' Prominent among these influences was choosing methods that previous developers had used in COS development, as discussed in the following section.

I basically followed what other people have done

Published and ongoing developers alike described looking at the literature to see how others had previously developed COS, and then deciding to use methods based on the

literature. Developers described feeling comfortable with methods that had been reported in the literature, particularly if the methods had been frequently used, as described by this developer who used the Delphi method:

> *I was aware of people who'd done similar types of work […] and I looked at, at their published work and they'd used the Delphi methodology and it seemed to generate the same sort of things, the same […] style of findings that we were looking to identify. So that's basically, that's basically why we arrived on that.* (O28)

Developers spoke of particular published COS that had influenced the methods they had chosen. For ongoing developers, this included HOME for eczema and Sinha et al.'s work in paediatric asthma. Only one developer reported looking at both published and ongoing work, and cited work ongoing in otitis media (O20). This developer remarked that their COS development group had '*gained experience from that*' and were trying to learn from the examples they saw. Of all available specific examples of COS, both published and ongoing developers talked about the OMERACT work as both the most common and strongest influence on the methods they chose. Developers regarded OMERACT as established leaders in the field of COS development:

> *So one thing we, err I really looked at and we all really looked at was the work that O-OMERACT has done, err they, they sort of seem to be leaders in the field, and their, their OMERACT handbook, well the 2.0 now I think […] set out as a step-by-step guideline, that was something that we took really to heart and something that we really based a lot of our work on and sort of tried to follow that, and sort of the con-consensus and having multiple phases and their, their approach […] we sort of really took that to heart.* (O32)

In contrast to developers who used previous work to inform their methods in a direct or 'off the peg' way as they '*didn't want to reinvent the wheel*' (P16), other developers (both published and ongoing) talked about using previously used methods as a starting point but then engaging in an evaluative decision making process about the suitability of previously used methods. For example, P5 described looking at previous COS work and then thinking about what it would mean to use the Delphi method that others had used previously in other clinical areas, and asking whether they could '*replicate that work*' in their area. P5 described using what had been done by others as a guide rather than a rule. P5 also said that they '*knew that other people had been fairly successful with this sort of approach [Delphi] and the rationale behind it is very convincing*'.

Describing an ongoing COS, O30 said that they '*basically followed what other people have done*,' but went on to describe how they adapted the methods, suggesting that they had thought more about their own specific circumstances for using the methods rather than simply applying the methods exactly as others had done previously:

> *We basically followed the eczema, um, core outcomes, […] adapted it a little bit in terms of the scale they used, how they, um, they progressed from one, er, round to another. You know they had like, er, criteria, predefined criteria, of consensus. Er, I think they had a lower threshold, and we had it higher because we thought that you know with so many outcomes we need to be more strict […] we should have more people, er er, agreeing on, er, that the outcome should go forwards from one stage to another.* (O30)

Ongoing developers also described looking at COS development guidance documents to inform their choices about methods. For example, O21 described using the guidance as a basis for their methods, but also building upon the guidance to:

> *Create the methods […] we had to sort of modify a little bit, since I started off thinking specifically about patient reported outcomes and wanted to make sure we captured um, err physician assessed outcomes […] sort of modified our searches and added on to it as we went.* (O21)

These ongoing developers mentioned the COMET Initiative as a particular source of guidance, as indicated by O26 who described looking at this work and then working out which method would best fit their resources. Although this developer referred to published work from COMET, he also highlighted the absence of guidelines about the best methods for COS development:

> *We consulted previous work done by people in our unit, err and also by published work from the comment, COMET err, err initiatives, like Paula Williamson's publications and so on, and then we tried to work out which is the best method that we can fit in within the timeline and the manpower that we have. 'Cause some of them are a bit more, you know, time consuming and there is no consensus, to my knowledge, on what are the best methods.* (O26)

One published developer, whose COS preceded the COMET Initiative, remarked that '*we went into the literature about evidence and literature about outcome, err and curious enough we didn't find COMET at the time, so I didn't see it*' (P12). Furthermore, another developer described finding the COMET paper outlining the issues to consider when developing a COS, only after starting to plan their COS work for their grant application (O22). Although O22 remarked that it would have been better to know about this paper much earlier on, it was not yet published when this developer was planning their application. This illustrates the difficulty of working in a developing field of research rather

than necessarily a lack of awareness of the existence of guidance or problems with their dissemination. Furthermore, although the paper referred to above outlines the issues to be considered in COS development, it is not a formal guideline document.

We took approaches that we were somewhat familiar with

Both published and ongoing developers described choosing methods for COS development that they had previous experience with. Developers spoke of general experience with methods rather than experience specific to COS development, as in this example where P14's experience with the Delphi method informed the team's choice of methods for COS development:

> *The methodology of the workshop was essentially built on the experience of running three other Delphi-style workshops dating back over the previous, ooh, 10-12 years um where I and my colleagues sort of refined the process of preparing statements that people could vote on, and refining those statements um so that the face-to-face time was as efficient as possible.* (P14)

One ongoing developer remarked that their previous experience with the method allowed them to design a '*good*' Delphi survey for their COS project (O31), indicating a belief that their experience had resulted in a better quality design. Referring to the selection of interviews as a method, O24 described that as a team they had general experience with interviewing patients '*assessing their views about a variety of things over the years*,' and *'it just seemed logical to extend it to assess the views and experiences of patients […] as well as making sure that whatever was being measured was, was relevant to patients*.' O24 did not give any indication for why interviewing might be better than other methods but said '*it seemed the obvious thing to do*.'

In contrast to the developers who spoke of having general experience with methods, a few published developers described having previous first-hand experience of using particular methods to develop COS and using the same methods again in subsequent COS development. One developer described being involved in an earlier COS project as a subject matter expert, and then '*running the show*' in the later work where the study team largely overlapped with the earlier study (P13). Similarly, P15 described how some of the study team had been involved in the development of a prior COS, so they had *'been through the process*' and '*it seemed to have worked well*.' Neither of these developers

showed any indication of critical evaluation, or adaptation of the methods, in light of their experiences.

I guess largely we just took their advice

Again, both published and ongoing developers described expert advice as another influence on their choice of methods, as indicated by P2 who referred to consulting with methodologists who had recommended using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) working group methodology, because it was perceived by his team to be a more robust method than 'opinion alone'. The GRADE criteria were developed as a means of applying a quality assessment to evidence. The GRADE system classifies the quality of evidence in one of four levels—high, moderate, low, and very low [140]. GRADE uses symbolic, and number/letter representation for quality of evidence and grades of recommendation [141], but still requires interpretation by the assessor . P2 said that the team '*largely just took the methodologists' advice*:

> *If you really want to in a […] scientifically robust […] process that you can argue is, is, somehow more than just your professional opinion, use this GRADE working group erm methodology. And so that was their recommendation and they walked us through it.* (P2)

Similarly, P5 described seeking advice from people who had more experience in the area of COS development, and '*found a rheumatologist who said, you know, that's what life was like for us 20 years ago but we figured a bunch of things out*.' P5 described the rheumatologist as a mentor and '*so every step we took he advised*':

> *It was very nice because he would say, well, you know, here are people who've done this. Here's something to look at, these are, you know, new ways to consider* your *question.* (P5)

P5 returned to this later in the interview and elaborated:

> *We talked through, well how might we do this, you know? Um what should I go read to learn more about this Delphi thing I never heard of? Um you know, who could provide some advice about how this might, you know, how this could play out for us? All of that.* (P5)

A few ongoing developers referred to '*expert advice*' from the COMET Initiative, and choosing methods designed and used by one of the COMET Initiative group members. One developer used particular methods because they '*had faith in the methodology because*

[*name of COMET Management Group member*] *had designed it and used it*' (O23). Similarly, O25 described being '*very much guided*' by a member of the COMET Initiative, '*obviously with extensive understanding of methodology in this area*'.

Interestingly, a few published developers framed advice about COS development as instruction. P4 described how one person from OMERACT '*told*' the team what they '*needed*' to do. Similarly, P8 described his group's uncertainty about whether a face to face meeting was needed and how their decision to conduct a meeting was very much influenced by OMERACT:

> *I said you know if you do a robust Delphi and whatever online, or remotely, if it's robust enough do you have to meet, or can you just class that as your consensus; which other people have done and you know seems not unreasonable. But their view from their OMERACT experience was, in fact their very words were if you don't get people in a room it will never get signed off. And so it was on their advice that we decided to call people together.* (P8)

One developer described having to make changes to their methods because of a journal's influence (P9). In this instance, the COS developers had the opportunity to schedule their consensus meeting at an international conference because '*most of the surveyees attend this meeting*' anyway. The COS developers had originally intended to ask individuals attending the meeting to vote on whether they would recommend the outcome measures, but their plan was '*derailed*' because the journal *'wouldn't publish this paper if we did a vote […] they said we do not publish recommendations*.' As such, P9 described how they had to *'calibrate the finality of the meeting so as not to derive recommendations, although that was my initial intent is to tr-… kind of provide a stronger guidance than I actually ended up writing*'. This developers' account suggests that they had to soften their recommendations in order to publish them. No other COS developer described similar experience but the influence of the journal in this individual's account is particularly interesting because of the impact it had on P9's choice of methods, and in turn, the potential impact this could have had on the outcomes included in this group's work, and the strength of, the final recommendations.

It was striking that only one developer described asking patients for their advice about the methods that they were using (O25). This COS team, who were developing a COS to be used in a paediatric population, explained taking their ideas to a patient organisation to gain feedback from young people about whether they felt they could be involved in a COS

project. O25 remarked that children and parents made a *'significant contribution'* to the *'actual methodology and delivery'* of the exercise.

If I'd had unlimited time and resources maybe we could have done that

Published developers described that limited funding meant that they could not bring people together for face to face meetings. For example, P9 described how '*fairly small*' budgets meant they were unable to hold a meeting and bring together people from all over the world. Another COS developer commented '*well number one we didn't have very much money, so even though maybe it would have been lovely to have had some sort of consensus conference or whatever, there was no way we were going to get this um degree of expertise all in one place*' (P5). However, P5 went on to explain that such a consensus conference would not have suited their patient population '*for whom that sort of non-synchronous, non-face-to-face communication is ideal because their lives are so very complicated*'. Therefore, while P5 would have liked to have had a consensus conference for clinicians, it *'wasn't financially possible […] if I'd had unlimited time and resources maybe that, we could have done that.*' As resources would not need to be unlimited to conduct a meeting, P5 is clearly speaking rhetorically here, perhaps in an attempt to justify her decisions. In contrast, P14 described being '*fortunate*' that a pharmaceutical company provided funding for their team's work, and as such they were able to hold a face to face meeting '*within the limits of the budget available*'. Although they received funding from a pharmaceutical company to hold a meeting, P14 remarked that the pharmaceutical company did not influence the choice of methods for COS development, as they '*were at arm's length from anything that Pharma might want to do about influencing our thinking*.'

Ongoing developers commented that as well as budgets, time limitations influenced the methods they chose. O27 described that their team had twelve months funding to develop the COS, and while they had considered conducting interviews with clinicians as an alternative to focus groups, they thought that time would not allow it. O27 referred to focus groups as an efficient way of bringing clinicians together, as well as a way of getting clinicians to be very open:

> *Focus groups are a really efficient way to bring clinicians together. It can be hard to get clinicians in a room um to do a focus group but this seemed to work best in our context. Um and we, we also thought a qualitative approach to garnering that information was really appropriate, 'cause we needed to give clinicians the opportunity to, to, to raise things that they thought were important rather than*

> *asking them a set of fixed questions um and ticking boxes. We had to be very open at the focus group stage to what might emerge.* (O27)

Likewise, O26 described having to work out, in the absence of any guidelines, which methods suited their resources best:

> *We tried to work out which is the best method that we can fit in within the timeline and the manpower that we have. 'Cause some of them are a bit more, you know, time consuming and there is no consensus, to my knowledge, on what are the best methods in, in this process.* (O26)

One COS developer described the impact of running out of time and funding on the choice of methods for COS development, which meant that the team could no longer use the methods they had originally decided upon:

> *The funding period ran out we were in the process of getting patients together to do focus group as well as the [clinical experts], but we just um ran out of funding and ran out of time to do… the survey was essentially a, essentially a first step to um, to determine how the [clinical experts] and how the researchers thought about the domains, and then the next step would have been actually having focus groups.* (O21)

Pragmatism I suppose is the ultimate answer

Some developers described basing their decisions about methods on pragmatism, that is, grounding decisions about choice of methods and their adaptation in practical and logical considerations. Published and ongoing developers repeatedly used phrases like it was '*a little bit pragmatic*' (P8) and '*pragmatism I suppose is the ultimate answer*' (O23). O22 talked about relying on '*logic*' about what methods would make sense in the context of their COS development work:

> *I have to say it was my own kind of logic that just kind of I used to develop the process. And, you know, just in writing this proposal for the grant just kind of thinking it through on my own in terms of what would, what would make sense.* (O22)

Some ongoing developers also talked about considering what methods to use but ultimately making '*arbitrary*' decisions about these. O19 described the process of thinking about different methods, and although they had an expert in conducting focus groups on their team (who was keen for them to use this particular method), they ultimately decided that interviews suited their objectives, '*to augment the list rather than find out the reasoning behind it*', better. O19 explained there were pros and cons to different methods

and the team were not aware of '*proof*' that one method is better than the other. He described their choice of methods as an '*arbitrary*' decision with '*no proof to back it up*,' but O19 did suggest a reason for their choice, it better suited their objectives, and not just personal whim, thereby contradicting the notion that it was arbitrary. O23 similarly described their decision making about methods as '*arbitrary*' and described how the team '*took a pragmatic view*' doing what '*seemed sensible*,' indicating some conflation of arbitrariness and pragmatism in COS developers' accounts.

Similar to O19 who referred to pros and cons of different methods, one other COS developer described how their team used a combination of methods, in a type of methodological triangulation, to deal with trade-offs of advantages and disadvantages of different methods (P15). They did this by combining a consensus meeting with their systematic review and Delphi polls, which P15 felt enabled them '*to make sure [they] hadn't missed anything*'.

Finally, two ongoing developers also mentioned that '*opportunity*' influenced what they did. In one example, there was an opportunity to develop a national survey and these developers decided to use that opportunity to ask about outcomes (O20). In the second example, the developers were invited to host a meeting at an international conference '*and it seemed like too good an opportunity to miss really to get their involvement, so, um, that consensus meeting was an open session at that conference*' (O23).

7.2.2 Challenges of COS development

All developers discussed the challenges they faced throughout the process of developing a COS. An overarching concern related to resources for COS development. This was already touched upon in the above section in relation to influences on methods chosen for COS development. The impact in relation to other challenges is discussed here. Generally, developers spoke about the difficulty of getting funding for COS development and the challenge of completing COS work in a timely manner. COS developers' accounts of the challenges conveyed an overall sense of compromise in how they went about COS work as a consequence of limited funding:

> *The most difficult part was um, well to combine in general, was to combine the resources we had with what err ideally should be done for developing a core outcome set. Because see, if you really want to do a systematic review of the*

> *outcomes err included in clinical trials, if you want to do some qualitative research err to have more the patient perspective, to take more the patient perspective into account, and if you want to do a Delphi study like, as we did, so to focus also on both quantitative and qualitative perspective, um it's err, I mean, you know, quite some resources are needed […]it's not easy to get money in err, in this field. So we had a limited amount of money, so we really had to match err our resources with err, um with what we ideally wanted to do.* (O31)

Specific examples of how COS developers compromised on methods in order to work within available resources are discussed throughout this section of the chapter.

You can only do as much as you can do if there are no guidelines

Published developers described being uncertain about the best way to develop a COS, but they did not directly attribute this to a lack of guidance in COS development. In contrast, ongoing developers explicitly pointed to the absence of guidance in COS development, suggesting that ongoing developers have become more methodologically aware as the area of COS development has established. Published developers had questions about what methods were available for COS development, which methods were most suitable in particular contexts and how to conduct a COS study generally. Both published and ongoing COS developers also talked about learning, over the course of their studies, how complicated COS development was; as in this instance where one developer said that his team was '*a little bit naïve*' at the outset and did not realise the challenges they would face until they faced them (O23). O23 went on to highlight problems his team had encountered with the Delphi method, such as access to participants, sample sizes, response rates, how to feedback results and the influence of the method of feedback, whether to retain or discard outcomes along the way, the weight to give to different stakeholders, and whether to combine different stakeholder groups or keep them as separate. O23 described these as '*all sort of mysteries*' that no one really understands, suggesting layers of complexity to the COS development process that were only discovered as the project progressed.

As noted above, ongoing developers referred specifically to an overall lack of methodological guidelines to develop COS: there is no '*determined methodology*' (O25) and '*there's no obvious guidelines about the definite ways of doing these things*' (O28). Ongoing developers talked about the absence of a '*fixed template*' and had doubts about whether such a template was even a possibility, as even if something had worked in the context of a COS in one condition '*there's no guarantee*' that it will work for another (O19).

COS developers also described a general lack of understanding among researchers generally of the process to develop COS '*correctly*', and doing the best they could in the absence of guidelines to develop a COS:

> *There is a bit of, er, a lack of, I don't know, understanding of, er, you know core outcomes process, and, um, why it's actually needed. And, er, you know, how it should be done correctly. I am not saying that what I've added was 100% correct in everything, but […] you can do as much as you can do if there are no guidelines.* (030)

Ongoing developers' accounts of the challenges of COS development were more specific than those of published developers. Only one published developer described the challenges specific to one method, the Delphi method, and simply stated that '*there's no standard, or even standardised way of using the Delphi process*' (P15). Ongoing developers described methodologically orientated challenges in some detail, as in the excerpt below where one COS developer described the challenges the team experienced in doing a systematic review of outcomes (O19). O19 was part of a group of researchers who considered themselves to be experts in conducting systematic reviews, and yet they still found the process of conducting a review in the context of COS development challenging:

> *Doing the review in itself was err, was very challenging, and particularly err tedious err because we realised that it, it wasn't really a quantitative type review, it's err, it's more um, you know, quant-qualitative/quantitative, semi-quantitative type review […] we, you know, wanted to look at the breadth and scope of outcomes which have been reported in trials and it was difficult to set thresholds: where do we stop? Do we stop at RCTs, do we stop at, you know, non- randomised comparative studies? Do we include case series? Err do we include commentaries or reviews? […] it was essent-essentially a new, new set of rules that we had to set for ourselves.* (019)

One developer who was working to produce a COS intended for use in both research and practice described how this added to the challenge. The team was conflicted about whether COS should be developed for both research and practice together, or developed for one specific context at a time (O22). O22 felt developing a COS for both research and practice was the right goal overall, and therefore developing the COS for use in both settings was the right approach to take. She went on to say that because they were successful in getting funding for this specific task, which she attributed to there being a need for consistency in the outcomes for both research and practice, this reinforced that this was indeed the right goal.

Other developers described being uncertain about how to define consensus, including the challenge of deciding cut-off points and the percentage of consensus to use, when others had defined consensus in various ways and there is '*no kind of consensus*' about how to define or achieve consensus (030).

Specific areas where developers suggested that guidance would be helpful are discussed further in section 7.2.3.

Even if you've engaged them for round one, whether they'll stay on for round two is another matter

Published and ongoing developers alike described challenges in retaining participants, as well as sampling the 'right' people in the first place:

> *Probably the biggest challenges are around um trying to make the broadest outreach and assure that um people who need to be in the room or need to know about will be invited to be in the room are there.* (P13)

> *The major challenges, you know, having all people that you need in place, you know, you have several people in, in the field that are in the driving position and, you know, you cannot always have them participating in […] war exercises, and the second is trying to um, you know, take into account all different opinions and all different approaches, and that is not always possible.* (P10)

COS developers discussed the management of conflicts of interest, as in this example where P15 described how *'in a small field like [name of disease], if we had excluded people who had developed instruments, many of the um best researchers in the area would be out.*' As a way of managing this challenge, this group asked instrument developers not to vote on their own instruments. P15 also described being transparent in their paper about who was included in the process.

The challenge of including multiple stakeholder groups was also prominent in both published and ongoing developers' accounts. There was agreement that despite the challenges it brought, multiple groups should be included '*from the beginning*' and '*very much integrated*' (P17), as doing so makes the COS '*ultimately stronger*' (O25). Nevertheless, this brought difficulties, particularly when the groups were brought together and O32 described how groups tended to be quite inflexible, only taking their own perspectives into account:

> *One of the problems that we saw is that sometimes some groups are little and narrow minded and they sort of have their own view, and they don't sort of go out of that specialism. So parents will be more focused on one thing, clinicians will be focused on another thing, where we sort of want them to, to sort of take the, the view of the other side into account as well.* (O32)

Developers also described challenges with low participant response rates, and mentioned strategies they employed to improve response rates:

> *The biggest challenge that we faced was err trying to get enough responses to the online survey. Despite [name of charity organisation] err help we, we struggled to get as many responses as we would have liked and we went to… We, we, we looked at the variety of ways around that including adverti-advertisements in the press and err, a-and other things.* (O24)

COS developers talked about having to send many reminders and '*chasing*' participants to respond. This was illustrated by O26 who described chasing every participant individually three times, which was '*quite a lot of chasing, quite a lot of trying, and the response rate was not great at all*'. When probed about why they might have experienced low levels of response, O26 suggested possible lack of interest in COS development or '*because of time commitment*.' O26 also alluded to why they might have struggled to engage industry representatives in particular, attributing this to an absence of pre-existing contact or links to industry representatives. Similarly, P4 highlighted the absence of a patient self-help group in one country as a reason for struggling to recruit patients in that country, in contrast to other countries where patient groups were established and recruitment was successful.

O32 also referred to difficulties with response, and felt that their team's efforts to increase the number of participants were successful. In part, they attributed this success not only to sending reminders, but to widening their pool and identifying '*interested*' people to take part after initial poor response to a general email via European societies. The inclusion of interested people is an attractive strategy with understandable rationale, but it might work against representativeness:

> *One problem is err, 'cause adopting the approach that we did we, we sent out online questionnaires, so emailing people, asking them, what do you think? Please answer these questions; please can you answer them by a specific date, and then sending many, many emails sort of reminding them yet again, could you please answer our questions? and so on. So that's, that's the one downside is that it does, it does take a while and not everyone was very responsive. Err you know, when researchers and people who are sort of leaders in the field get 50 or 100 emails a*

> *day, when they're told, could you please answer a questionnaire, even if it's really going to take two minutes, they're not super willing to, to help out quite often. So it took, it took some time but in the end err we had a fairly good response rate, we thought.... So I think that was the one sort of key issue of finding people who are actually interested in it and not just, you know, emailing anyone who was in the field.* (O32)

Other developers also described low response rates as a consequence of participant burden, such as this example pertaining to patients:

> *They found the technical side of err completing the online questionnaire very frustrating and eventually gave up. Quite, quite a, quite a few, I would say maybe another 20 to 30 actually gave up after trying.* (O19)

Some developers highlighted how the platform they used for their surveys had influenced response rates, as in this example where one COS developer described wanting to send their online survey out, as a link within an email, via a clinical society. The society '*weren't keen*' but then after '*their initial reluctance that they didn't want to get involved in a questionnaire survey*' they reached a '*compromise'* to send it out as an attachment that people would complete and send back:

> *That was a real faff. And from the 200 odd people on the mailing list we got about 50 responses. Um, I think a load more people would have done if they'd just had to click a link […] I wouldn't have done that (laughs) […] even though it's only 20 seconds more.* (P8)

Ironically, when asked why the clinical society was reluctant to send out the link, P8 responded that as he recalled '*from conversations at the time, they didn't want to overburden people*.'

COS developers highlighted the importance of '*who'* it is that is being included as participants. One COS developer expressed concern about the difficulty of achieving diversity in sampling patient participants and partners for COS development and how not achieving this could lead to overlooking outcomes that were important for less advantaged patients (P17). She recognised that having a diverse sample is a resource issue because it takes more time to recruit a diverse group, and these participants may need extra support to allow them to participate.

> *I feel really strongly about this. It's about this issue of um who we're including as patient research partners, because the problem is, and this happened a lot in [name of clinical trial], um 'cause initially one patient partner was invited from each European country with, with their [specialist] to sit in and talk about the domains, um and they, they, they were all white, they all spoke English and they were a very particular kind of patient. And it, it worries me a lot that the people who are being involved in this way on a sort of patient research partner level, but probably as*

participants generally, um don't represent that diversity, and so we're missing potentially priorities, you know, a different set of priorities, and we've, we've got to find a way of including those. (P17)

Developers also described maintaining participation as a challenge, as explained by this developer who highlighted potential consequences, in terms of imbalances in stakeholder groups:

The worry has always been um lack of um engagement from clinicians err and, and secondly, even if you've engaged them for round one whether they'll stay on for round two err is another matter. So there's a large imbalance in the two stakeholder groups at the moment, we've got about 110 patients for the patient group and we've got only 50 for err, for the clinician group. (O19)

Having a patient participate in COS development would be challenging

Published developers spoke of including patients as participants in COS development as complicated and challenging, and pointed to this as a reason for not including patients in their studies. Published developers were also of the view that it was somewhat unusual to seek the opinions of patients in COS development when they undertook their COS work. This was in contrast to ongoing COS developers, of whom all but one had included patients in COS development. The impact of not including patients in COS development was described by P17. A COS already existed when P17's core set was developed, and had been developed with clinicians only. She described how trying to get the '*patient core set to sit alongside the professional one…was always going to be a big challenge*.' She remarked that other developers who are integrating different stakeholder opinions from the outset would not encounter this problem.

Published COS developers were usually of the opinion that patients would lack understanding of the concept of outcomes and of research more generally, and that this would make it difficult for patients to be able to select important or appropriate outcomes.

I am not sure I would involve patients in a final decision about endpoints. In my experience, even trained researchers sometimes have difficulty grasping the concepts about biases in endpoints. So, I believe patients would be similarly challenged, if not more so, and their involvement could make deciding upon the appropriate/valid/unbiased endpoints and primary endpoint more difficult. So I'm not sure on that one. (P7)

We would have struggled um to bring a patient in, I think they would have needed some coaching and some background information on what the meeting's um topic

> *was and what the discussions were […] you know, we had thematic discussions around research methodology, we had a nice controlled trial versus cohort trial versus, you know, and, and how different variables could be structured in those trials and I, I think having a patient participate in that would be challenging if it wasn't someone who had a knowledge base, you know, to enable them um, you know, to contribute.* (P13)

Irrespective of these anticipated challenges, P13 commented that '*having patients participate and weigh in on meaningful outcomes is, is a critical piece*' of COS development. This sentiment was echoed by other COS developers. For example, a developer who attributed their lack of patient participation to two major barriers, time and budget, described not realising the importance of including patients until the work was done, which suggests that they might not have truly considered patient participation at the outset:

> *We felt like we missed an entire dimension around [name of disease] by um not including patients at every level of the stakeholder engagement process. And it was actually only by doing the, this project that we realised that.* (P6)

P17 was one of the few published COS developers to describe experienced, rather than anticipated, challenges of involving patients in COS development. She described how patients found it hard to prioritise outcomes, and how the team had found it difficult to convey the realities of research and clinical trials and the need to narrow the outcomes down to a feasible number. P4 also described how patients rated everything as important, and as a direct result the team had to modify their methods to answer a slightly different question, '*what should be included into the core set*' rather than which outcomes are important. Having anticipated that this could have been a problem, P5 described how they spent a long time thinking about phrasing the question, emphasising the importance of anchoring it in the real world '*within the realm of possibility presently*' and relating the question to an intervention. P5 also anticipated that when patients were asked what they would change about having their condition, they would answer '*well I won't have [name of disease] anymore…that kind of almost magical thinking*.' P5 seems to be attributing a lack of realism to patients' view about what is important to measure, but it is entirely realistic for patients to want to get rid of their condition. This idea will be explored further in the Discussion.

P17 described taking time to give patients an '*explanation in a way that will make them feel confident enough then to contribute*'. Patient participation was described as '*resource heavy*' but for this published COS developer (P17) the benefits clearly outweighed the

challenges. She explained how it was important to ensure that patients could recognise that '*their voice was there because it was different from what they [patients] would normally hear in their clinical setting*'. This highlights that patients can use different language to clinicians to talk about outcomes. In this COS developer's experience, '*approaching people face to face*' meant that they could '*talk to them and explain*' the study and as such '*they totally get it and they want to be involved*.' P17 also highlighted the importance of appropriate phrasing and terminology:

> *We're doing something right here because they're responding to this, we've got the right phrasing, um and I think that's a really important aspect […] in that Delphi we took a lot of care around the phrasing so that um the patients wouldn't be sort of alienated by the terminology used. Um and at the same time not alienating the professionals I suppose, but they, they could understand, we sort of had a, a double layer, so there was sort of um the more technical term that you might get in the literature alongside um the wording that came from the qual-qualitative data. Um but err definitely more resource heavy.* (P17)

O19 echoed other developers in reflecting on how patients found it difficult to understand the nature of what they were being asked to do as participants, adding that because COS development is '*fairly abstract'*, '*there's always going to be a little bit of bias towards the more educated people*.' This is an important issue for COS development that will be considered further in the Discussion section of this chapter. There were some indications that one source of patients' difficulties with understanding COS development may be the methods used to involve them. O19 utilised an online Delphi in contrast to P17 whose face to face approach seemed to be more successful. O25 also remarked that patients and parents did not understand the idea of core outcome sets until it was explained face to face, then '*they got it*'. Interestingly, O19 described these as '*mini-challenges rather than significant challenges*', suggesting that these issues were relatively easy to overcome. This raises questions about whether COS developers need to take more responsibility for making COS accessible to patients, a point picked up by O32 who questioned '*how well does the person who is explaining it actually explain to the parents?*' O32 proposed providing more information and possibly writing participant information down, however more information does not always lead to better information or a more informed participant.

Similarly, O29 commented on how a patient dropped out of the Delphi process who did not fully understand about outcomes in the context of a trial. Again, this could reflect a failure with the researchers' explanation. It also suggests that it is a complex task to portray the

relevance of COS, and their value, to both research and clinical practice. This raises an important question about how COS developers can make COS relevant to patients and help them to be aware of the relevance of COS to the clinic:

> *Some of the patients, um certainly I know one, when it came to the Delphi, um at that point I think he grasped that it was about outcomes in trials, and I don't think he was that clear about that before, even though we'd gone into some discussion about it and um obviously our information had all that in it. Um and then he didn't complete the second round of the Delphi and he said, well what I'm more interested in is outcomes in the clinic, and then he didn't complete the second, second round of it. So yeah, I think it's very important from, you know, to be very clear um about what it's about.* (O29)

This particular COS developer questioned whether such problems arose because researchers did not communicate about COS well enough, or because '*for some patients it just is beyond their grasp*' (O29), adding that '*there will be different levels of understanding amongst different patients*.'

Difficulties with prioritising outcomes was not exclusive to patients, as COS developers described clinicians also finding '*it hard to discriminate between outcomes*' (O20). COS developers explained how they tried to overcome this particular challenge, for example P5 described being optimistic that after one survey they would have the core outcomes, but as it turned out, that was not the case. '*Because the first way didn't work'*, the team introduced an unplanned survey to '*give people the ability to really weigh their choices*' by giving participants 100 points to distribute amongst the outcomes, giving the most points for the most important outcomes. This was a post-hoc decision that the team made once they discovered that people rated everything as important, but P5 said that '*it didn't really help'* either. In contrast, P8 described asking participants to rank their top 3 outcomes, as the team had anticipated that '*all the outcomes would be important to someone*'. P8 remarked that this had been '*one of the most important methodological choices*' they made:

> *We wanted to see whether choosing your top three requires a little bit more thought than ticking a list, which may not require more thought but requires different thought, you know lists you are prioritising. So we thought well let's just see how, um, this method can help discriminate between outcomes that, yeah are, you know are important, or outcomes that are the most important.* (P8)

In contrast to patients, only one COS developer described a lack of understanding of COS among clinicians. O32 remarked that if clinicians did not know what was being asked of them or if they had choices that they did not fully understand, *'very often they are going to*

just pick one, they're going to guess, and then that becomes an issue.' This might also suggest unease with working in an area that is fundamentally opinion based. O32 also suggested that clinicians might be reluctant to '*admit it*' because '*doctors are not always very happy with saying, I don't know*.' Similarly, only one COS developer described challenges with researchers' ability to grasp concepts about outcomes (P7).

Ethical boundaries

Like ongoing COS developers who did not include patients but anticipated that resources would be a challenge of including patient participants, ongoing COS developers who did include patients emphasised how the lack of resources to support patient participants was a significant challenge. COS developers referred to financial, that is funding, and human resources as the main resource needs for involving patients. COS developers saw the resource intensive nature of involving patients as linked to the ethics approvals that were required in order for patients to participate in research. Because of the time and money associated with ethics approval processes, some COS developers decided not to include patients in COS development:

> *One was time, so this was a one-year contract that we had from [name of funding body] um and so in order to involve patients in our process, we felt like to do it well we would want to get um IRB approval um to en-, engage focus groups of patients, and felt like we didn't have sufficient time to do that unfortunately, or budget to sort of pay out incentives for that kind of thing. And so in the end we felt very comfortable at the time um involving patient advocates and not patients themselves.* (P6)

O31 elaborated about why ethics applications were particularly resource intensive, describing how countries outside of their own had ethical approval procedures that they '*were not prepared for*.' As a result of the unexpected procedures, these developers were unable to involve patients from all countries that they wanted to:

> *It was not easy to involve patients in the Delphi survey […] we wanted to do it with patients from different countries, but it was not easy to reach the amount because err in some countries we had to um go through some ethical approval procedures that we were not prepared err for […] in the UK for example, so we were not able to involve err patients from the UK, from, for this reason, because we tried to do it through patients' organisations in the UK but err indeed we had um to apply [for ethics approval] and err we did not have err time and err resources to go through all this process. So in the end we had to make very practical decisions err not to involve patients from some countries err where we really wanted to involve patients and we um… so we did it only from other countries where there were not these err ethical boundaries let's say.* (O31)

O27 succinctly described the ethics procedures associated with stakeholder research in different countries as 'the bureaucratic hurdles of international approvals'. Some COS developers also differentiated between stakeholder involvement as participants in research and '*stakeholder work which isn't classed as research*'. O27 referred to COS projects which had been described as the latter, '*and when that happens you don't have the um err the large workload of the um ethics processes to deal with*.' One COS developer described a range of challenging administrative tasks associated with ethical approval for involving patients, and commented on how these were much more taxing in the US than in the UK, where this COS developer was based. O29 also described the US process as bureaucratic and excessively complicated:

> *The US was a, was a totally different story, that was extremely, it was extremely slow and laborious process to get um ethical approval for the study, um I think there were some concerns because we were going to be discussing some issues with patients that um, mean we had to have a psychologist that was available that, that patients could refer to if there were concerns about um low mood. Um and, and just um, you know, the bureaucracy there is […] something else compared to here, you know, there's still a lot of hoops to jump through, and it just took us ages to get it off the ground.* (O29)

Kids can't always tell us what they're feeling

Published and ongoing developers perceived particular challenges with involving children in COS development. Remarking about involving children as participants, one COS developer (P1) who had not actually sought children's input, commented that it was '*hard to ask them'*, and added:

> *So I think we kind of know what they would say. We kind of know what they would say; that they wouldn't want any pain, so I think pain would be the only outcome they would consider*'.

One ongoing COS developer who had not involved patients similarly anticipated that children's involvement would be challenging due to their limited language abilities, '*which makes it a little bit more difficult because they can't always tell us what they're feeling*' (O32).

To be, or not to be (international)?

Both published and ongoing developers talked about the challenges of doing COS development internationally, particularly the linguistic challenges that global participation

entailed and the need to translate concepts and questionnaires. Some COS developers described how, having anticipated this challenge, they did not include international participants, while others confronted this challenge and involved international participants:

> *'Cause that was the, that was one of the other difficulties, 'cause we had err parents in five or six different countries, so we had different language err questionnaires.* (O32)

> *We kept our study within the UK, um, I would have liked to have done it more internationally, um, but um access to participants is an issue, and, um, funding and various other things. We did our Delphi on paper so posting things internationally was going to be tricky, and then there's language issues and all this other stuff.* (O23)

O32 went on to describe further challenges relating to participants' understanding when questionnaires need to be translated into different languages:

> *It's a drawback because the question is, is the questionnaire really the same in every language? Is the err, do parents understand the same thing from the questionnaire in every country?* (O32)

Alternatively, COS developers described an international COS that only included English speaking participants, which engendered different challenges particularly for including patient participants:

> *The main challenge was, er, the language thing, obviously that was, er, in English. So we, for example, okay for clinicians for example, the most of the clinician and the researchers, and of course journal editors, they do speak English. But then the problem was to actually find patients who would be willing to take the questionnaire, all three stages, fill it in, and have a proper understanding of English.* (O30)

COS developers also spoke of the logistical challenges of organising an international meeting, and the challenge of getting the balance between what is ideal and what is pragmatic. O25 suggested that a kind of interim 'clinician COS' could be developed with international clinicians, and then there could be country specific work with patients followed by a consensus meeting to determine the final overall COS for each individual country. O25 thought this approach would be more suitable than running an international Delphi with patients across many different countries simultaneously:

> *The barriers in terms of language particular, um and reaching a conse- a-and I think a consensus meeting is important, and I think strategically to get patients to fly from around the world to a consensus meeting is not really realistic, and therefore I appreciate that that's, wh-what I'm suggesting is not an ideal answer, but I think it's pragmatic.* (O25)

O27 described a similar scenario whereby limited resources might mean that an international approach is not possible even if desired. COS developers engaged in a trade-off between the best way of doing things within the resources available:

> *It's best if these things are international actually, but I think we have to think very carefully about what) we are doing the research for and whose care we might be improving. Um we also need to think very carefully about the resources we've got available um and that, that includes financial resources to do a piece of work err as well as the human resources we've got to do a piece of work, which of course are interconnected, but it might be that sometimes we just have to do work um in, in a single country setting for those reasons, and that, that, that's okay. Um sometimes that's a pragmatic approach to take and it makes good sense to do something in, in one, in one place.* (O27)

One COS developer differentiated between a COS that involved international participants and a COS that was developed to produce international recommendations (P14). He had included participants from different countries but this was because there happened to be expertise in these particular countries. It was not because they were trying to include international representation to produce international recommendations:

> *If you look at the geographical origin of the people in the workshop […] All the rest were from mainland Europe, UK and North America. No South Americans, no Africans, um so some people would say it's not an international consensus. But we didn't really (.) seek to present it as such. We were um aiming to produce an output from an expert group. When in some other workshops you want to maybe generate outputs that provide convincing guidance throughout the world - in that case you want to be as international as possible - that becomes er frequently quite unwieldy because of language barriers in, in discussion, in the workshop situations and (.) different levels of development of understanding in the particular field. Um, for instance, I think the awareness of investigators involved in [name of disease] trials in Asia is significantly um less sophisticated as to methodology compared to people in, especially, Europe and North America. So we ended up with a bias um to those areas for a good reason, because that's where the expertise was, that's where the critical thinking had been occurring.* (P14)

The issue of whether COS should be international or not was described as an '*unanswered question*' by one COS developer (O25). Another COS developer saw the answer as very much linked to the overall goal of the work:

> *I don't know if the, if the step would be to have all stakeholders, you know, from all different countries in the same room, or whether there should be parallel processes in different countries that potentially reach very different conclusions but that should be sort of compared. And so I, I think that that would need to be considered in terms of what the method would look like. But the focus of our paper was, the goal was not to sort of get international perspective, it was to get a US perspective.* (P6)

If no one uses it, you cannot really say that the development of your COS was successful

Ongoing COS developers described challenges with getting their COS protocols published, with one even referring to the process as '*painful*' (O19). In all instances, this was linked to reviewers' and editors' perceived lack of understanding or knowledge about COS, the development process and the importance thereof:

> *It wasn't really easy to publish it […] not many people are actually experienced, are experienced in, er, core outcomes, and they are not very experienced in how core outcomes should be, you know the e-Delphi kind of process. Because (coughs) I was trying to publish in, er, another, er, journal before the, the one that, which I published it at the end. And from their reviewer point of view you would get you know very funny sort of questions and critiques, which, which just wouldn't make sense to someone who you know, um, who would appreciate how difficult a process it is.* (O30)

Some ongoing COS developers expressed concern that the time taken to develop COS might mean that the COS is '*slightly obsolete*' when the COS is published and ready for use (O19), as '*things change*' such as '*new developments, new interventions come into the equation*'. COS developers felt that given that COS development '*takes a little while to do these things, staying at the forefront of the evidence can be hard as well too*' (O21).

Interestingly, only one published COS developer described difficulty when publishing the findings of their COS work. P8 remarked that he had received '*disparaging*' comments from a couple of high impact journals, and had been asked '*this is a nice idea but is it really science?*' (P8). This COS developer went on to say '*the people were a bit disparaging about the idea that we, um, just asked people what items were important.*' P8 described two further challenges associated with publishing their meeting report. The first that '*peer review takes forever*', and the second that the reviewers did not like one of the scales they had decided on, and did not '*agree with the consensus we have made*.'

Published developers described uptake and implementation of the COS as a challenge of COS development more frequently than ongoing developers, with some published COS developers describing implementation of the COS as the '*biggest challenge*.' One explanation of this difference is that published COS developers are more aware of the challenges associated with this stage of the process, while ongoing COS developers, with their focus on more immediate concerns, were deferring some of their efforts to promote implementation to a later point in the process:

> *I don't see it too much as a challenge now, although see, again if you have limited resources, not that you can have a very broad implementation plan, but this might be a challenge after some years when our core outcome set will be developed and we will be seeing maybe that people don't use it err too much. (*031)

Nevertheless, despite not regarding implementation as an immediate challenge, this COS developer went on to describe the importance of implementation from an early phase, commenting that if no one uses the COS once it is developed it can hardly be regarded as successful:

> *You can err have the best methodology in developing a core outcome set but err, err if at the end of the day the, no one uses it, you cannot really say that your, the development of your core outcome set was successful. Implementation is not an easy step but it is really worth it to spend err time and energy on that because err again um having a core set which is not used by the clinical com-com-community that err it's the target to might be er r not very successful.* (031)

He emphasised the importance of adopting '*implementation strategies*', and went on to describe a dissemination strategy to aid implementation:

> *So we have reached consensus on outcome domains, and um yeah, the first step of course we made a publication on this, but then err, err we are trying to work err on newsletters err of professional bodies for example, so to implement more our um, our message. We have presented the results of our study in err three, four different congresses and err we are going to present in even more err, err during this year, so to present the results of these different conferences in different err, err also with different audience possibly.* (O31)

Other COS developers described instances where COS implementation had been unsuccessful and they did not know how to resolve the situation. Linked to this, one COS developer described a lack of openness by gatekeepers to the uptake of their COS:

> *So the FDA is, right now, studies, erm, are being planned and conducted on systemic treatments for [name of disease] and the FDA just doesn't accept our core outcome set. They favour an outcome that is not included in the core outcome as a primary endpoint, and they just refuse to acknowledge our work. And we wrote them, I think four or five months ago, about this issue and we never got a reply*. (P4)

A similar challenge was echoed by another COS developer who described how the existing leading authority for recommendations and guidelines in their field felt '*threatened*' by their COS work (P9). The COS group had planned to do further work to involve patients, but according to P9 the leading authority were so '*miffed*' by this COS work that they created a '*political heat storm*.' As a result, the COS developers decided not to do this further work as they did not want to '*antagonise*' the committee further. P9 relayed how the two groups

held discussions and the leading authority agreed to be more open to some of the discussion that arose from their COS work. He attributed this to the leading authority wanting to remain in control as the standard for recommendations, rather than have the COS group '*derail'* them.

Published developers were aware that the process of COS development could influence its implementation. Describing this as '*social and political'* aspects of COS development, one COS developer explained that who is involved in the process of deciding what is core can influence the later uptake of the COS (O31). Another COS developer highlighted similar issues:

> *Probably the biggest challenges are around um trying to make the broadest outreach and assure that um people who need to be in the room or need to know about will be invited to be in the room are there. Um I've seen really good consensus initiatives go bad um because err one individual was insulted that they were not invited to participate, um literally. That was um, err a, a group I didn't participate in but I, I, I know a woman who moderated a consensus um, err initiative recently and, you know, again she got together five or six really good people, put them behind a closed door, have them do all kinds of literature review and worked very hard to put out a document, and when they rolled it out there it was met with a lot of resistance among um the community of [clinical experts]; it was a very, it was a big challenge.* (P13)

When asked what the solution was to involving all necessary stakeholders to avoid problems with implementation, P13 answered *'Magic wand?*' which suggests that she saw implementation and uptake as something that was virtually impossible to get right. In some respects COS could be regarded as proposing that no one group has authority and that all opinion is equally important or valid. This can be extremely challenging for some groups. P13 went on to offer some solutions, including the planning committee '*knowing who to bring into the room, and, not keep it confined to people who are similar in thought to you, right?*' This COS developer also described what she felt were successes with her COS project, such as having '*an organisation that is perceived as non-biased*' to help oversee or facilitate the process and to overcome '*the feelings of inclusive or exclusiveness*.'

Ongoing and published developers talked about people's lack of understanding of COS, their importance and their purpose, as barriers to the uptake of COS recommendations. COS developers described an apparent lack of knowledge about what COS are for, along with a concern that COS might stifle other outcomes of interest that should be included in research:

> *The biggest challenge, erm, maybe to make, erm, people that are important in a field but not directly involved aware of the importance and of the definition of the development of a core set, and also of the importance to then apply to it without necessary having to feel restricted in their choices of like primary outcomes or whatever.* (P4)

Similarly, P12 talked about a general lack of awareness of COS as a challenge to implementation, and saw this as COMET's responsibility to resolve:

> *COMET's problem that when you are in an area and so knowledged about it, and you are working full time in that, you are speaking an internal language which a lot of people [may] not know exactly what you mean outside, and that is one of the major implementation problems. You need to have this out so people can understand what we are talking about.* (P12)

P11 described frustration of speaking at a meeting where people disregarded the recommended outcomes as '*irrelevant*,' and went on to say '*It's almost like people weren't even listening to what you had to say*.' He added that when people have a '*basics of understanding*' about outcomes and the importance of core outcome sets it is like *'speaking the same language*,' which can help with uptake and implementation of COS.

O32 described that implementation of a COS might be challenging when COS recommendations are based on a small number of experts, asking '*Do we trust this group as a sort of oversight committee over a larger group of experts, or do we need more people involved in this?*' O32 also suggested that the status and trustworthiness of the people leading the COS, as well as the numbers of participants included, might also influence uptake:

> *Will there not be criticism that, okay well you did this but, you know, who are you? Why, why should we trust your core outcome set? You only asked a hundred people, and so on and so forth.* (O32)

O28 asked similar questions, again highlighting the importance of process issues such as the selection of participants in the subsequent uptake of the COS:

> *How do you convey to people that who you've chosen is the, are the right people? Um you know, how do you know or how can you err how do you know, and how can you demonstrate if you do know, that the people you've chosen are the right people to be commenting on it? So I think that's probably the major thing when we've sort of been talking about this amongst you know when we've been presenting the work locally that's the thing that's perhaps generated the most discussion.* (O28)

7.2.3 Guidelines for COS development

As already highlighted in section 7.2.2, COS developers emphasised the absence of guidelines to develop COS, and suggested areas where they felt specific guidance would have helped them or would be helpful for future COS developers. They also suggested areas of research where they felt there were methodological questions that need answering:

> *It's really helpful to be developing these guidelines because at the time that I was doing this there wasn't really, you know, I, I felt a bit like a pioneer in a way, you know, trying to decide which was the best way to go about this. Um and I think there are some really important discussions that need to be had.* (P17)

One has to be careful that any kind of description of what gold standard is, is also tempered by a reality check

COS developers described the need for guidance about how to develop a COS. Published COS developers spoke of a '*best*' methodology, implying that they thought there could be one right method for COS development. P1 suggested that guidance is needed about '*which technique should be chosen*', and P13 similarly spoke of a need for '*some semblance of guidelines […] for what is an optimal methodology for getting to a consensus*,' or '*a preferred methodology*'. P2 remarked that guidance for the novice investigator or for clinicians who do not do research full time, is needed, and went on to say that it would be helpful if the entire process was laid out and COS developers '*could be aided by a concrete set of steps to go through*.' P2 likened this '*set of steps*' to CONSORT, adding that '*people need to be educated*'. This COS developer also suggested that guidance should be a '*prescribed process*', suggesting that guidelines should be set down as rules, while commenting '*I don't know if you would, if a journal could mandate that you followed a particular order of steps.*' P2's turn of phrase here suggests that this COS developer had concerns about how such prescriptive guidelines could be enforced. Nevertheless he also suggested that a concrete set of steps to go through would '*make it a more robust set of outcomes in the end, that people actually believe and care about*,' adding that guidelines about COS methodology could impact on the uptake of that COS, by allowing people to decide if a COS has been rigorously developed:

> *People could say wow they really did not stick to what is considered a robust methodology for creating a core outcome set, I really have no faith in the final product or the converse, they did everything they could given the resources that*

> *they had and I do think that future investigators should lean on that paper. Or not, you know.* (P2)

Despite suggesting the need for guidance to describe how a COS should be developed, COS developers offered few suggestions as to what the COS development process should entail. Only one COS developer made a suggestion that a combination of methods was most appropriate, but did not explicitly say what that combination should be. This COS developer did, however, say that '*a systematic review is very important, where there's enough literature to do a reasonable systematic review*' and also said that he preferred the Delphi poll as it means '*you can actually get, if you're going to do a consensus […] because it focuses attention'* (P15).

Other COS developers similarly said that a flow diagram and set of steps like CONSORT would be helpful for COS developers. CONSORT offers a standardised way for authors to report clinical trials, and it comprises a 25-item checklist and flow diagram for trialists to follow [11]. P11 remarked that '*what I really think is helpful is you can talk about all the theory you want but what people really want is a "how to" guide […] a step by step march through it.*' Guidance needs to *'lay it out for people as to that process they march through*' because *'people don't think about these things at all.*'

> *Now, look, I'm a pragmatist, okay. I'm all for laying out the conceptual foundation but at some point the rubber's got to hit the road, and that's what we're not doing and needs to get done. It's to lay out this here's the CONSORT type check list for developing outcome measures as end points in clinical trials. And I would even go beyond clinical trials – I would say clinical research in general. It doesn't have to be in the setting of a randomised trial.* (P11)

P12 also compared guidelines for COS development as being similar to CONSORT, adding that establishing guidelines for COS development would be a challenge because of the need to cover '*so many areas'*, but this could be overcome by identifying '*common rules, […] for all type of medical areas.*' If methodological guidelines were disease specific, '*people outside might not understand it and say it's not for us, we jump out*,' so there is a need for '*common rules'* to be established. Likewise, in contrast to those published COS developers who described the need for guidance about a best method, one published COS developer described the need for guidance about *'how it can be done'* rather than a single, gold standard, way of developing a COS (P4). This COS developer suggested that a protocol template would be helpful for people to follow, as well as '*a group of advisors or more experienced people in this field who would be available to consult new groups*.'

Ongoing COS developers also felt that guidance was needed for how to develop a COS, but unlike published COS developers, they acknowledged that this could not be '*a one size fits all approach*' due to the '*nature'* of COS development (O19). O19 spoke about a need for guidance to help researchers '*think for themselves*' about how their individual circumstances influence how they develop their COS for their different diseases. Although he did not feel that a '*one set fits all*' guideline for how to develop a COS was possible, he did suggest that '*what you can do is have some guidelines on how to be flexible'* as '*essentially there's no fixed template to do this*.' Likewise, O23 suggested that there is no fixed method for COS development, and instead proposed that COS guidance should be developed to prompt COS developers to think very carefully about their data sources. O23 described how his COS had involved qualitative interviews with patients, but later found out that extensive qualitative research had already been published in their field. He therefore felt this qualitative work had been done unnecessarily. He added that such interviews could be useful in other clinical areas where such work has not been done previously. These views further support the idea that COS development is not a fixed process and guidance should reflect that. O27 further elaborated this point, commenting on the need for COS teams to make '*considered decisions about the methods to use along the way,*' and illustrated this with an example of methods they considered for their context:

> *We discussed whether to do focus groups or interviews with clinicians and working out what works best in a timeframe, what works best for tho-those people. All those have to be thought about in some detail um in order to get them right.* (O27)

In direct contrast to P2 who suggested that COS guidance should be prescriptive, O27 said '*I think again it's maybe about not being prescriptive and ensuring that people do have the chance to think about their own contexts, and work in accordance with their contexts really, and, to kind of respect local knowledge, I think, as well*.' Rather than being prescriptive, '*being aware of some of the possible hurdles, that is really important.*' Nevertheless, O27 also highlighted the importance of guidance being practical and taking into account resources issues.

One COS developer highlighted the importance of not only developing guidance for COS development, but of COS developers knowing that guidance exists and where they can find it (O22). When prompted about whether there were any particular guidelines that she

would have found helpful, O22 described how she had not learnt about the COMET papers until after the group had started their COS work (the COMET papers were published around the time that this COS team started their work). She described discovering '*things*' from reading the COMET papers that their team had not done. She saw COMET as very much part of this process of disseminating COS development guidelines:

> *When I started this whole thing I did come across COMET, but I don't know if those guideline papers were out there yet or… Like I don't know why I didn't find them the first year I started this […] I was quite far into it before I came across some of those other papers, then I was like, oh crap, I should have done this or that, kind of thing, you know? So I think just getting your stuff out there more so that when people are starting it that they find it.* (O22)

Ongoing COS developers commented that guidance around how to develop a COS would also have a positive impact on dissemination and uptake of the COS. For them, guidelines about methods would help make publishing COS work easier as they could refer to guidelines to justify their methodological decisions:

> *Some of the very helpful guidelines and papers that are coming out, like that would be something that, you know, if I could refer, you know, if there was a statement that I could refer back to, because I think a lot of journals don't necessarily know about this methodology yet either, and if I had a reference that I could point to that would be very helpful.* (O22)

Likewise, O30 said:

> *You have the guidelines, and you can tell referees, er, you know, we actually did what the guidelines suggested so it, it gets you know more straightforward.* (O30)

The particular COS developer quoted above (O22) highlighted the need for ethics approval as an example of something that they had not thought about during the development of the COS. She commented that it would be necessary to include advice about ethics approval in COS development guidelines. In the case of this group's COS, the issue of ethical approval only arose at the publication stage when the team started to question '*how should ethics work with this?*'.

Methodological mysteries

As touched on in section 7.2.2 above, it was almost always ongoing, rather than published developers who, suggested specific methods where guidance would be helpful. Two COS developers suggested that guidance '*on how to run discussion meetings*' would be useful. P14, the only published COS developer to suggest specific methodological guidance, described how they did not come across any existing guidelines, and '*20 years or so that*

I've been involved in chairing discussion meetings, I've learnt on the job, and I don't think that's necessarily the best way it's good to be able to benefit from other people's experience.' He elaborated that guidance should include *'how best to organise a workshop like this to try and ensure that you use the expertise and time of the individuals actually face to face as efficiently as possible*.' P14 also suggested that planning and preliminary work is vital to the success of the meeting, as '*ultimately you've got to have something that provokes people to think about a question and then to make a judgment*'. Another COS developer (O26) remarked on the absence of guidelines for consensus meetings specifically, commenting that there was variability in the literature about how meetings were conducted; some hold an informal chat whilst others have involved more formal voting. O26 remarked that there was no consensus '*on how to actually do it'*:

> *The functionality of the consensus meeting is very, very vague. In the, in the literature, there is absolutely no consensus there about how to do it and what are the objectives and the aims of having this consensus meeting, so that err something that needs to be looked at a bit further in detail.* (O26)

COS developers suggested specific areas of Delphi where guidance is lacking or further research is required. These included numbers of people for a Delphi, including number of patients (O29), and '*how do you know, and how can you demonstrate if you do know, that the people you've chosen are the right people?*' (O28) and how to increase response rates – '*how do you motivate people to answer?*' (O32). O32 went on to ask *'if you're motivating people to answer them, is that not biased?*' Furthermore, he remarked that he did not know if these questions would ever be answered:

> *So that was sort of the question where we sort of thought about it, okay we can tell the head of the department, tell the ten people in your department that we've emailed to answer it, but then we know that their answers will probably be similar to their boss's because they have a similar approach to everything, and then you bias your own results. So that's, that's something that's an issue, but I don't know if there's a, if there's an answer to it, I don't know if there's a good way to try and increase that.* (O32)

O23 asked similar questions about sample size and populations in Delphi surveys, but also had questions about the *'the internationalness of the population'* and the additional complexities that international participation brings, such as access to participants, funding and language issues. O23 listed multiple research questions about Delphi that need to be answered:

> *I don't think anyone understands the implication of, um, response rates in Delphi studies, because they are different to standard surveys, so I don't think anyone really understands that. […] within Delphi, I don't think anyone understands the best type of feedback to give people, um, the impact of what that feedback, um, you know of how that feedback affects responses. Um er, the issues around, um, whether you discard or retain outcomes, or whether you keep the whole lot in, um, until the end of the Delphi process, I don't think that's been resolved at all. Um what else had problems? Um, I guess the relative weight of different, um er, of different stakeholders, and do you keep them equal, do you weight one more heavily than the other. Um, do you, um, combine different stakeholders, different, er, results at different stages, do you merge it all at the end, um, yeah these are all sort of mysteries to be honest with you. I could go on and on.* (O23)

O23 also referred to the systematic reviews that are often done prior to the Delphi exercise, and the need to have methods in place that allow '*appropriate categorisation*' of outcomes prior to the Delphi exercise. He described that domain categorisation during their systematic review of outcomes was *'a subjective process'* that *'if you don't get that stage right then your Delphi just kind of doesn't matter.'*

The amount of involvement (and participation) that you have from patients will differ depending on the clinical context

COS developers made some reference to including patients as participants in COS development as an area requiring guidance. This included the '*number of patients*' that should be involved, as well as the '*amount of involvement that you have from patients*' (O29). This COS developer remarked that '*I guess that for every clinical context that will probably differ*' which suggests a belief that any guidance in this area could not provide a definitive answer to these questions.

O27 highlighted the importance of considering the resources that will be available to COS developers when thinking about producing guidance. She elaborated that '*patient involvement in research – a lot of groups won't have the resources to do that properly*,' and so guidance needs to be realistic about what is do-able with limited resources. P17 was the only developer to distinguish between patient participation in contributing data and patient involvement as research partners. She described both as resource issues, and the difficulties associated with representing patient diversity with limited resources. P17 remarked on the importance of making sure that the outcome priorities of different stakeholder groups are not missed, and on the need for guidance on how to represent diversity in both patient research partners and patient participants. P17 also acknowledged

the difficulties in producing such guidance '*we're probably not going to be able to give guidance'* on how best to represent patient diversity, but added '*going forwards we need to think about this, I think that would be a really important contribution*'.

How to report your findings

A small number of COS developers, both published and ongoing, commented that guidance on '*how to report your findings*' was currently lacking and would be helpful. One COS developer (O26) who described using the '*paper by Sinha et al*' to write their manuscript, when prompted about the usefulness of this reporting guidance, said that this was '*definitely helpful, and especially when you compare it to previously done previously published papers, you see that quite a few of the papers are missing very important points*.' This COS developer elaborated that reporting guidelines are important for '*transparency of how you've done your Delphi survey.*' Despite using the Sinha checklist, he still felt that reporting guidelines were necessary ''*cause I think that it's only very few papers that actually highlight, you know, best practice in reporting findings for the consensus.*'

7.3 Discussion

As the first in-depth description of COS developers' choice of methodological approach, including the factors that have informed how researchers have developed COS, this study has furthered our understanding of COS development processes and enabled me to identify priority areas for future methodological research. A summary of the main findings, including these priorities, is presented below.

7.3.1 Main findings

COS development is an emerging field and in this respect contrasts markedly with the area for which COS are being developed – clinical trials. There are many rules and a good measure of consensus about how to conduct clinical trials, and it is now a highly directed field with well-developed guidance to support the process [55, 56]. By contrast, COS development is an emerging field without shared assumptions, and COS developers are trying to bring order to the outcomes that are measured in trials in a specific condition, where currently there is considerable inconsistency in what is being measured and reported. The absence of a knowledge base for COS development and the lack of

consensus about the best way to develop a COS underpin methodological decision making processes. COS developers discussed a variety of factors that influenced their choice of methods, which included accounts in the literature of previously used COS development methods, expert advice, their general experience with methods and the resources available. Some COS developers described uncritical application of methods, for example using the methods they were familiar with rather than working out which methods would best suit their purpose and context. Other COS developers demonstrated more critical evaluation of methods, evidenced in the adaptation of methods before applying them. Experience with methods was a common influence on choice of methods, but this experience was general and not COS specific. Furthermore, COS developers who had previous COS experience did not demonstrate any critical appraisal of their methods before applying them to subsequent COS.

Questions about resources, namely time and money constraints, also influenced COS developers' choices of approach. COS developers emphasised a need to compromise in how they went about COS development as a consequence of limited resources, and saw this as one of their major challenges. While small budgets and time limitations are likely to be genuine constraining factors, it is possible that in some ways these serve as justification of choices about methods rather than explanation. For example, COS developers who received generous funding still talked about their projects being compromised due to funding and time. Time has been widely accepted as a critical barrier to research, but could also be suggestive of the task in hand as being seen as lower priority than other activities [136].

Interestingly, there were examples of advice being framed as instruction. While it is possible that this advice was given as instruction, it could be that COS developers saw certain well known figures as leaders in the field of COS development and therefore followed their advice without questioning it. The notion of instruction also contradicts the opinion of ongoing developers that there is no single way (or right way) to develop a COS. It is also possible that these COS developers were seeking to avoid accountability for their decisions, or mitigating responsibility.

COS developers frequently cited OMERACT as an example of previous work that had informed their decisions about methods, and the OMERACT team as providers of advice.

While OMERACT members have years of experience in COS development, their experience is limited predominantly to the field of rheumatology. OMERACT have developed the OMERACT filter 2.0 as a template for rheumatology COS development [27, 142], and there have been recent similar efforts in the field of dermatology [143]. Condition specific guidelines certainly have their place, but as one COS developer highlighted, if guidance is disease specific then others might be reluctant to use it. COS developers saw the COMET Initiative as having a role in guideline development and dissemination in the field of COS development.

Ongoing COS developers referred to the available guidance in this area, albeit limited, as being useful in deciding what methods to use. Of note, ongoing developers were keen to point out that there is no consensus about an optimal method for COS development, so they were still required to make their own decisions about which methods to use, whether this be to suit their resources or their circumstances. Ongoing COS developers somewhat self-critically referred to having made arbitrary decisions about methods, but they seemed to be conflating this with the need for decision making to be influenced by pragmatic considerations. Developers who described their choice of methods as pragmatic conveyed an element of logic and thinking through their methodological choices, albeit within resource constraints. This is reflective of an emerging area of research where guidelines on which to base decisions are currently limited. The absence of guidance in this area is discussed further, later in this chapter.

In contrast to published COS developers, ongoing COS developers described more detailed methodological uncertainties pertaining to the process, particularly systematic reviews of outcomes, defining consensus and uncertainties about the Delphi method. These more detailed descriptions of methodological uncertainties might reflect an overall increased awareness of the complexities of COS development, through the work of groups such as the COMET Initiative [40, 144]. One published COS developer referred to their work as preceding the '*methodology movement*' implying that researchers today are more methodologically aware. One other explanation is that description of more detailed methodological uncertainties could also be a matter of recall, as the ongoing developers were still very much involved in that work at the time of interview, and these issues may therefore be at the forefront of their minds.

One of the biggest challenges experienced by COS developers was around participant selection, access, and retention. COS developers also expressed concern about reflecting multiple stakeholder opinions and integrating different opinions into the process. Health researchers have long looked to systems and frameworks in an attempt to understand human endeavours and interactions, and complex adaptive systems are one such system [145]. Complexity here relates to the importance of interactions of parts, in relation to the establishment of the whole [145]. COS development could be regarded as a complex adaptive system, that is the collection of individual groups or agents with freedom to act in ways that are not always predictable yet whose actions are interconnected [146]. Many of the concepts of complexity apply to COS development, such as neither the system nor its external environment being constant, individuals within a system being independent decision makers, and having to manage uncertainty and paradox within the system. For example, individuals within a Delphi study are making an independent decision about the outcomes that are important, but when feedback is given they become interconnected to the group as they can change their scores based on the collective. These exist not necessarily as problems that can be solved, but as ways of understanding COS development. Furthermore, because the agents or groups within a complex system can change, the complex system in itself can adapt over time [147]. This adaptation can occur through learning or experience. The activity of COS development in its entirety can, at a higher level, also be considered a complex adaptive system. Through the sort of experience and learning such as indicated in some developers' accounts, this complex system is likely to adapt in time. With different groups developing COS in different ways, their experiences might lead to changes in the ways COS are developed, or even to more standardised ways of developing COS as more evidence about methods becomes available (i.e. through guidance), thus creating an adaptive system.

Incorporating diversity within stakeholder groups was also an issue, and is an important challenge to address as it may influence the outcomes ultimately defined as core. Involving patients in the process of COS development was a particular concern for developers and an area of COS development where developers felt that guidance was needed. Published developers particularly perceived that involving patients in COS development would be challenging, and as a consequence most had not included patients. However, it is worth considering that this might be retrospective justification for not having considered patient participation explicitly at the time of development. Published developers also described

that including patient participants in research was not commonplace when they undertook their COS work, which might corroborate this suggestion. COS developers also anticipated difficulties with involving children in COS development. This is an area that requires further work to identify specific challenges with involving children, and to look for exemplar work where children have been successfully involved and from which COS developers might usefully learn.

COS developers suggested that patients had trouble understanding the nature of what they were being asked to do, but also questioned whether the problem was really that COS development had not been explained sufficiently rather than understanding being an insurmountable issue for patients. One developer suggested that providing more information to patients might help with patient understanding. However, it is important to highlight that more information is not always better information and might not be helpful at all. More information could even be detrimental and lead to patients disengaging because it leaves them feeling overwhelmed or confused. One developer also attributed a lack of realism to patients' views about what it important, referring to patients not wanting the disease anymore. While it is true that outcomes for trials need to be realistic and measurable, it is entirely realistic for patients to want to be rid of their condition. This arguably demonstrates the vital importance of patient participation in COS development, as a reminder for researchers not to lose sight of key research goals, even if these currently seem out of reach.

Interestingly, only one COS developer highlighted the issue of researchers' ability to grasp concepts about rating the importance of outcomes. Additionally, developers spoke of how both patients and clinicians struggle with rating outcomes. Again, one could argue that it is more about how the question is framed, so what it is that participants are asked to do, and the importance of differentiating between important outcomes and core outcomes that should be measured and reported in every trial, rather than participants' inability to discriminate between outcomes. One also needs to question whether understanding is so important. If participants do not understand what is being asked of them, then one needs to question what this means in relation to participants' ratings of outcome. COS developers did not exclude clinicians or researchers from the process based on their perceived level of understanding, yet some did exclude patient participation because of it. However, most ongoing COS developers are now including patients in the process.

Increasing patient participation in current COS development provides an opportunity for further research with patient participants themselves to find out about their experience of being included in COS development. We can then learn from this to inform guidance in the area of patient participation in COS development. Interestingly, patient involvement did not feature heavily in developers' accounts in these interviews, with only one developer making a clear distinction between patient participation in contributing data and patient involvement as research partners. The involvement of patients as research partners is an area of COS development that requires further attention.

Some COS developers proposed including patient advocates (as participants in COS) as a solution to challenges of involving patient participants. Here, developers were referring to including individuals who can 'represent' the views of the wider patient community. While research has suggested that training patients to actively participate in the research process may increase confidence and ability to participate in research [148], COS developers need to be cautious about this and question the 'representativeness' of trained patient advocates. Patient advocates tend to be of a higher socioeconomic status [149], which could have implications for the outcomes they think are important. Ives et al describe the professionalisation paradox whereby they suggest that training patients undermines the very purpose of their involvement to bring a lay perspective [150], and propose that researchers risk being drawn into a paradox that will render patient participation and involvement ineffective, even tokenistic. Patient advocates likely have a place in COS development, for example as research partners, but probably not at the expense of involving patient participants providing personal perspectives. This is an area of COS development that requires further exploration.

Interviews with COS developers also raised questions about whether the inclusion of patients in COS development is participation in research or consultation. The distinction between involvement and participation is arguably not so clear in COS development, as patients in a Delphi exercise or consensus meeting are in a sense being consulted as experts about what outcomes to include based on their knowledge and experience of a health condition. Furthermore, this links to the above point about professionalisation, and some have argued that consultation should not require the professionalisation of the patient to provide their experience of living with a disease or as a service user [150]. This issue was also raised in relation to ethics requirements, and is therefore an important

consideration for future COS developers. COS developers expressed that ethics requirements are currently unclear, or developers were lacking knowledge of the requirements, and there is a need for guidance around what ethics approvals are required. One developer felt that if COS development is classed as consultation rather than research then ethics approval would not be necessary. The Health Research Authority offers a decision tool to determine whether a study requires Research Ethics Committee review (http://www.hra-decisiontools.org.uk/ethics/). This tool suggests that if the aim of the study is to generate generalisable knowledge, which COS are, then it is research rather than consultation, and therefore an ethics review is required. Of note, if COS development claims the status of research rather than consultation, ethics consideration would be required for all participants, not just patients.

A particularly prominent question among developers was whether COS should be developed internationally. This links to the intended reach of the COS recommendations. If the COS is intended to be used globally then this has implications for how COS are developed, who is involved in that process and the resources required. Considerations about efficiency and heterogeneity arise with global development, as well as generalisability. However, heterogeneity can be as great within countries as between so this should not serve as a barrier to internationally developed COS. In a letter to an editor in reference to the international HOME for eczema COS, it was pointed out that for a disease with global impact there was limited representation of non-western participants from countries where the societal burden of the disease is high [151]. If a COS is developed to have international applicability then there is an issue of inclusivity that needs to be addressed. The question of international representation in COS development is one that requires further research and insight.

In this study, COS developers expressed concerns about the translation of questionnaires when multiple language participants are included, which could, in part be addressed by the literature on translation of instruments. Limited work has compared different techniques for translation of instruments, including back translation, forward translation and use of multiple methods [152]. The World Health Organisation (WHO) recommend an approach that uses both forward and back translation, input from experts, pretesting with target populations and methods of suggesting alternative wording and expressions [153].

International Society for Pharmacoeconomics and Outcomes Research (ISPOR) have also developed principles of good practice for the translation and cultural adoption process for patient reported outcome measures [154]. Researchers can decide which method suits their purpose, but research could be done to see whether the same techniques could be applied to questionnaires and surveys, such as those used in the Delphi method, about outcomes in the context of COS development.

Implementation of a COS was seen as a key challenge by COS developers in this study. They described a lack of openness by gatekeepers to adopting COS, a general lack of understanding within the research community about how COS function and their benefits, and the importance of process issues such as the selection of participants, in influencing uptake. The link between appropriate stakeholder participation in development and subsequent implementation highlights the importance of considering implementation of the COS earlier on in the development process. Developers felt that COMET has a role in raising general awareness about COS. However, there is a shared responsibility in raising awareness and gaining interest in COS, as developers have a role in achieving this within their clinical and research communities.

From their accounts, it seemed that some COS developers adopted what could be regarded as a rather passive approach to implementation. Some did not seem to think much beyond publishing the COS, while others did think about implementation but saw this as something to address at a later stage of the COS development process. Developers' accounts raise a complexity about impact and implementation of COS. However, the success of implementation is not just a product of the quality of developers work and efforts to publicise the COS and promote impact. The readiness of the target community is of utmost importance to its success, as demonstrated by some of the opposition faced from individuals and organisations in these developers' accounts. Precedence should be given to devise strategies to convey the importance of COS and encourage relevant communities to implement them. It is therefore critical to discuss implementation in any future COS development guidelines.

Integral to the development and implementation of COS are questions about how far work in this area is valued by developers themselves and also by the wider community. For COS work to be successful it needs to be valued, yet the developers' accounts pointed to ways

that COS activity is currently undervalued. This could be linked to the challenges of promoting understanding about COS, and to the unease that developers expressed around opinion based nature of COS work. With developers reporting peer reviewer comments such as '*is it really science?*' (P8) and encountering disparagement because they *'just asked people what items were important*', thought needs to be given to how the value attributed to COS work can be communicated and enhanced. This will be considered in the Discussion in Chapter 8.

The absence of guidance in COS development, and the prominence of uncertainties, dominated COS developers' accounts. There was an apparent distinction between published and ongoing developers' attitudes towards the nature of guidance in this area, from one of seeking prescribed ways of doing COS work from published developers, to one of seeking guidance that promoted thought and good practice by ongoing developers. There was recognition amongst ongoing COS developers that context was very important, and that what worked for one COS development might not necessarily work for another. Developers saw a need for guidelines that identify commonalities across disease areas, and aid researchers to think about their own contexts and circumstances and how this influences what they do, as well as being aware of potential challenges and hurdles. Whilst a gold standard is aspirational, guidelines need to be practical, realistic and resource friendly. COS developers also commented that guidance on how to report COS work was currently lacking and would aid future COS developers. COS developers understood that answers to methodological questions and uncertainties may not be immediately forthcoming, but felt that raising awareness of important areas and potential pitfalls in itself would be a significant contribution. It was particularly felt that guidance around the systematic review process, conduct of Delphi, and conduct of consensus meetings, were high priority. Guidelines for consensus meetings may well be found in other contexts or specialities, such as clinical guideline development and many of the issues with using this method may not be unique to COS development. This should be explored before work is undertaken to avoid unnecessary duplication. As the most frequently used methods in current COS development, there is an impetus to conduct necessary research to answer questions about best practices for these methods and provide guidance to COS developers.

7.3.2 Robustness of the study

Strengths of the study

The strength of this study lies with the conduct and analysis of qualitative interviews to provide insight into, and a first in-depth account of, experience of COS development. The use of semi-structured interviews meant that I was able to cover specific topics of interest, whilst remaining flexible to adjust questions and change the direction of each interview as they were taking place. The survey responses in Chapter 5 had to be taken at face value, whereas responses to the interview questions could be clarified and expanded in the moment.

The iterative nature of the analysis enabled me to incorporate new topics or questions into subsequent interviews. Furthermore, my use of framework analysis maintained a link between the original data and the conclusions drawn from it, thus adding to the trustworthiness of the conclusions drawn from the data.

Due to the broad nature of the topic guide used, this study contributed to understanding about the complete process of COS development, which in turn has identified aspects where guidance or research will help to improve future COS developers' experience and improve standards of COS development. Finally, by including the most up to date COS studies, that is, those published after 2010, as well as ongoing COS, the findings are relevant to present COS development, and indeed for future studies. In terms of the concept of catalytic validity, which refers to the usefulness of research and its potential for real world impact [137], the findings from this study point to priority areas for future methodological research, which in turn has the potential for a direct impact on future practice in COS development.

Limitations of the study

This study provides insight into the experiences of only those COS developers who agreed to be interviewed. Furthermore, I identified ongoing COS developers through the COMET database, meaning that COS developers had already been in contact with the COMET Initiative to register their work in the COMET database. This sample could comprise a particular subset of COS developers and limit the transferability of the findings. Nonetheless, COS developers did represent a broad range of disease categories and COS

characteristics (methods used, stakeholders involved), supporting the transferability of the findings. It is, however, possible that this group of COS developers might have different experiences and views to those not interviewed.

Furthermore, the existence of established contact with COMET could have influenced their responses, for example developers' might have been more likely to provide idealised responses. As discussed in Chapter 6, it is possible that my association with the COMET Initiative may well have influenced the data. Lastly, the length and breadth of the interview topic guide meant that it was not always possible to cover every topic with every participant.

7.3.3 Summary

This study set out to generate a detailed description of COS developers' choice of methodological approach, including the factors that have informed the ways in which researchers have developed COS, and to identify priority areas for future methodological research. It is clear that the COS developers who were interviewed found the process of COS development to be a challenging process, in part due to the nature of COS development being an emerging field of research, but also in part to not always considering important methodological details from the outset. This study provides the first insight into ongoing COS developers' methodological decision making and their experiences in COS development in the current COS environment. The findings raise important questions and highlight areas that future research should focus on to solve 'methodological mysteries' and provide guidance for COS developers. Guidance needs to promote awareness of important issues; encourage COS developers to think about their own contexts and circumstances, and enable COS developers to make decisions about methods that best suit their needs and resources.

The findings of this study can be combined with the systematic review findings (Chapter 3 and 4) and survey of published COS developers (Chapter 5) to establish a research agenda and priorities for COS development guidance. These will be discussed in Chapter 8.

Chapter 8: General discussion

8.1 Summary of findings

COS are increasingly recognised as important for the design, conduct and reporting of randomised trials, systematic reviews and other forms of research. Accumulating work in the area of COS development has identified the need for general guidance on the development of COS. My principal aim in this thesis was to explore what influences COS developers' choice of methodology and approach in COS development, and identify priorities for guidance and further research in this area. Several conclusions can be drawn from the work within this thesis. A summary of each chapter is provided below.

When I started working on this thesis, COS were scattered across the health literature. A systematic review of COS was needed to bring these resources together in one place. This systematic review also enabled me to describe the methods developers had used previously, and also to help inform the later work in this thesis. In Chapter 2 I described the development of a search strategy to identify COS. This was particularly challenging due to the variability in the use of free text terms and index terms on reports of COS development, the absence of MeSH headings in MEDLINE or index terms in other bibliographic databases for identifying COS papers, and inconsistent categorisation across different databases. The search strategy that I developed for identifying COS was designed to be highly sensitive, but as a consequence a large number of records were retrieved, mostly irrelevant, and had to be assessed for suitability.

In the systematic review of COS described in Chapter 3, I identified 198 published studies (250 reports) that described the development of COS for measurement and reporting in clinical trials. COS development is frequently described across a series of publications. This was the first time COS from all medical areas had been brought together in a single place. This is particularly important for COS to be successfully implemented as they need to be easily accessible to researchers and other key groups. The systematic review highlighted great variability in the ways that COS had been developed, particularly the methods used and the stakeholders included as participants in the process. Furthermore, key aspects of

the process were frequently not reported. A striking aspect of the results was the infrequency with which patient participants had been included in the development of COS. While the systematic review provided information about the range of methods that had been used to develop COS historically, in order to inform current practice it was important to know more about the methods that were being used presently. At the end of 2014, the COMET database included a total of 57 projects listed as ongoing COS studies. The most common combination of methods planned or used by these ongoing studies was a systematic review, together with a Delphi survey, followed by a consensus meeting. In Chapter 4 I therefore described additional data, from the systematic review, about studies that had used these specific methods. This chapter further demonstrated the variability in COS methodology and corroborated the conclusion of the main systematic review, reported in Chapter 3, that there is a need to improve the standards of reporting in these studies. The variability demonstrated by the review, and the absence of an accepted gold standard method for COS development, meant that further work was needed to explore this variation. I anticipated that it would be helpful if we understood developers' rationale behind their methodological decisions; exploring choices of methodology and subsequent experience of those methods seemed important to inform how developers might be able to do this work better in the future.

I used a mixed methods approach to further explore COS development, drawing on qualitative methods and an online web-based survey. I conducted an online survey with published developers identified in the systematic review of COS. The survey provided quantifiable information about developers' experience, as well as the first insight into developers' methodological decision making about choices of methods. It also provided insight into developers' experiences of doing this type of work, including the challenges that they encountered. The majority of survey respondents (73%) felt that there is a need for methodological guidance or research to inform future activity to develop COS, including: stakeholder involvement, patient involvement in particular; choice of methodology, and consensus formation.

I undertook semi-structured qualitative interviews with a purposive sample of developers of published, as well as ongoing studies, and described the methods for this in Chapter 6. I decided to focus the qualitative study on more recent (published) and ongoing COS studies (unpublished studies), as my aim was to inform future guidelines for developing COS, and

so it was important to try and understand the current situation. One of the main findings, described in Chapter 7, was the variety of influences on developers' choice of methods, which included the previous literature on COS development, expert advice, developers' own experience with methods and the resources available to developers. The absence of guidance in COS development, and the prominence of uncertainties, dominated developers' accounts. By exploring developers' experiences inductively, I was able to provide the first in-depth description of developers' choice of methodological approach. It has furthered our methodological understanding of COS development processes and enabled me to identify priority areas for future methodological research and guidance, which will be the main focus of the rest of this chapter.

8.2 Comparison of COS developer survey and interviews

Developers' choices about methods were a focus in both the survey and the semi-structured interviews. In both studies, developers presented multiple reasons for the methods they used. The influences on the choice of methods for COS development were largely consistent across both studies, with the following influences featuring prominently in both: previous work; expert advice; experience with methods, and resources. As might be anticipated, the interviews provided more detail than the survey about these influences and how they shaped COS development. For example, it was apparent from the interviews that developers did not always apply critical thinking about the methods used previously in the literature, or as advised by experts, before using them. This has important implications when the methods that have been used previously are not the most suitable methods. Furthermore, the previous use of methods has been in other contexts (e.g. to develop a COS for a different disease). This knowledge would not have been gleaned from the survey alone.

Methodological decision making about COS development, as developers described it in their interviews, was informed by their previous experience of the methods generally, rather than specifically using the methods in the context of COS development. Although the response option for the question in the survey that related to experience referred to experience with methods for COS development, developers' comments to accompany this response suggested that it was their general experience with methods, rather than knowledge of the methods within the specific context of COS development, that influenced their decisions about the methods they used. This is unsurprising given that COS

development is an emerging field. Written comments accompanying the survey response option 'Suited our situation and circumstances' were more analogous to the accounts of interviewed developers who described basing their decisions about methods on pragmatism and opportunity.

All of the developers who were interviewed described the challenges of COS development and many of these matched the challenges that developers reported in the survey, such as involving multiple stakeholders, defining consensus, implementation of the core set and working within available resources. The interview findings went beyond the short survey responses by pinpointing the detrimental impact of some of these challenges, such as the need to compromise methods due to resource limitations. Survey respondents identified 'How to measure core outcomes' as a challenge, but this was not highlighted by developers in the interview study. Rather than seeing the findings as contradictory and having to give precedence to one or the other set of findings in order to reach a conclusion, I considered how my methods might have influenced these findings. One likely explanation is that I highlighted to the interviewees that the focus of my interest was on identifying what to measure, rather than how to measure, at the start of each interview. It was necessary to explain the emphasis of the interviews in order to focus developers' accounts, but in doing so I might have limited developers' responses to exclude their experiences about this equally important part of COS development. By not limiting the focus of the survey, these views were still captured.

Some significant challenges were discussed in the interviews that did not feature strongly in the survey, including difficulties with ethical application procedures and requirements, and the complexities of international COS development. The former challenge was spontaneously mentioned by one developer at interview, and then incorporated in subsequent interviews. The latter challenge was an area that I had anticipated and included in the topic guide. Interviewees suggested that if a COS is developed globally then it has implications for how COS are developed, who is involved in the process and the resources required. Considerations about efficiency of process and generalisability of the results arise with global development. The interview study highlighted that the question of international representation in COS development is one that requires further research and understanding. One of the limitations of the survey was that I could not prompt for topics. However, the question about challenges included an open response to list up to five

challenges. Both the interview and survey therefore helped to give a comprehensive picture of the challenges involved in COS development.

Only one survey respondent highlighted an absence of COS development guidance as a challenge. Sixteen survey respondents explicitly reported that they did not think there were any areas where guidance or research would benefit future activity to develop COS. However, 92 areas for guidance or research to inform future activity were listed by the remaining 57 survey respondents. 'How to measure' was the most frequently suggested area that surveyed developers wanted guidance on. How to measure core outcomes did not feature as an area for guidance in the interview study, but, as mentioned above, this is likely due to the emphasis being placed elsewhere during the interviews. A collaboration between COSMIN and COMET has recently resulted in the development of a guideline on how to select outcome measurement instruments for outcomes included in a COS [52].

The other areas where surveyed and interviewed developers felt guidance or further research were needed were consistent, particularly stakeholder involvement, choice of methods, consensus methods and implementation. The interviews generated much more specific information to help in identifying priorities for future research, such as the mechanics of the Delphi process (e.g. how to feedback results to participants, whether to retain or discard outcomes between rounds, whether to combine different stakeholder groups) beyond simply defining consensus. The interviews also highlighted the need for guidelines about how to report COS studies. Although this was not mentioned by survey respondents, it is a gap supported by the systematic review of COS that highlighted poor reporting of COS studies.

The interviews raised some important issues that were not seen in the survey. For example, developers' described problems around participants understanding the nature of what they were being asked to do, but also questioned whether the problem was really that COS development had not been explained sufficiently rather than understanding being an insurmountable issue. Understanding of COS by the wider community was also raised as an issue by interviewees, which had implications for the uptake of the COS. Furthermore, these findings raise the possibility that difficulties with understanding about COS might be linked to the value – or lack of value - that people attribute to COS. Some interviewed developers did not seem to value the COS work in itself, but as useful or necessary to

complete in order to achieve a different aim. Also relevant here was the unease that some developers expressed in their interviews about working with opinion based data. Surveyed developers similarly described a lack of data to support expert opinion as a challenge, as well as being the most frequently cited limitation of COS work. Due to the nature of the survey method, it was not possible to explore this issue further. However, by combining the survey and interview data we are able to glean more insight into this important issue of value; that developers' perceived that an implication of working with opinion data might be a subsequent lack of value being attributed to the COS.

Another interesting question raised by the interview study was about whether patient participation in COS development is participation in research or consultation in research. This question applies to all stakeholders who might participate in COS development. The distinction between involvement and participation is arguably not so clear in COS development, as participants in a Delphi exercise or consensus meeting are in a sense being consulted as experts about what outcomes to include based on their knowledge and experience of a health condition. As discussed in Chapter 7, The Health Research Authority offers a decision tool to determine whether a study requires Research Ethics Committee review (http://www.hra-decisiontools.org.uk/ethics/). If the aim of a study is to generate generalisable knowledge, then it is regarded as research rather than consultation, and an ethics review is required. If COS development claims the status of research rather than consultation, ethics review would be required for all stakeholder participation, not just patients.

8.3 Strengths and limitations

The strengths and limitations in relation to each study have been discussed in detail in the relevant chapters. This section will focus on the overall strengths and limitations of the work presented in this thesis.

COS development is an emerging field of research methodology. The work undertaken in this thesis has contributed to knowledge and has the potential to improve methodology for COS development. The identification and accumulation of COS to provide the evidence base for the COMET database is already proving to be a valuable resource for developers. It is important to acknowledge that the systematic review work provided the sample

population for the survey, and the published COS developer population for the interviews. The difficulty around identifying COS work in the published literature has been discussed extensively, but it is worth noting here that issues around searching could mean that relevant papers were missed, and therefore the populations in the proceeding survey and interviews were limited by this. The same can be said about the ongoing (unpublished) population that I drew from for the interviews, as these were developers already in contact with COMET. Other developers, not identified in the systematic review or in contact with COMET, could not have been included in this work.

The mixed methods approach undertaken here to explore developers' experiences of COS development is the first of its kind. The mixed methods approach allowed me to draw on the strengths of both methodologies and to provide a broader perspective on the overall issues. The interviews provided nuances that could not be captured in the survey alone. One limitation of the inclusion of open-ended responses in the survey, and the qualitative interviews, is the risk of idealised responses. That is, developers might have been inclined to say or write what they thought I wanted to hear, specifically when prompted during the interviews. There might also have been a difference in how developers articulated their choices in the survey and interviews, and how they made those choices in practice. This could work either way; those that described making uncritical COS development choices may have been more critical in practice than they alluded to and the reverse could also be true.

8.4 Implications

The thesis has identified several issues that are now being addressed by the COMET Initiative. These are discussed throughout section 8.4.

8.4.1 Reporting guideline for studies developing COS

The poor reporting of COS in the systematic review of COS, coupled with interviewed developers who expressed the absence of guidance on how to report findings, have confirmed the need for a reporting guideline for studies developing COS. Interviewed developers suggested that the credibility and success of a COS depends on the use of sound methodology in its development, and clear and transparent reporting of the processes adopted are required to demonstrate this. There is an increasing number of COS being

developed which makes this reporting guideline a priority, and to date no formal reporting guideline exists. A COMET research project, led by Jamie Kirkham (University of Liverpool), is now underway to develop a reporting guideline, along with an explanatory document to accompany the guideline [155]. The Core Outcome Set-STAndards for Reporting (COS-STAR) is expected to be completed in 2016. This work follows the approach proposed by EQUATOR [82]. It builds on a preliminary checklist of issues to consider when developing a COS that was based mainly on experience of the COMET Management Group members [38].

The survey also highlighted the importance of defining the scope of the COS; even if a COS has been developed rigorously it might not fit the required scope of the user. Scope needs to be well defined in reports of COS development to allow users to decide whether a COS matches their required scope. Failure to do so could result in the COS not being used, COS being used inappropriately or duplication of work to develop a similar COS for a specific scope. This is therefore an important criterion that should be included in reporting standards and any reporting guidelines for these types of studies to allow users to make this kind of decision. Items specific to clearly defining the scope of the COS have been included in the preliminary checklist of reporting items, to be considered for inclusion in the reporting guideline.

8.4.2 Quality assessment of COS studies

There is currently no way of determining or assessing the quality of COS studies and this work has corroborated the need for such an instrument. The survey has demonstrated that developers and users are making their own informal quality assessments and applying some criteria as to what they think makes a COS of good quality, such as patient involvement, and the level of methodological detail provided. However, without a standard means of assessing quality, individual judgements can result in different estimates of quality. In view of the continuing increase in the number of COS being developed, the development of a tool to determine or asses quality is a priority in COS research. As part of the COMET Management Group we have plans to develop a quality assessment tool. Discussions are underway about how we can best approach this and how one would define quality in the context of COS development. While the outcomes that are included in a COS cannot be subject to quality assessment, it is possible and desirable to assess

methodological quality of the process to develop a COS. Quality would therefore relate to the design of the study and the methods used, not the outcomes that are being recommended for measurement and reporting. We need to understand the methodological issues before we can make judgements about quality, so the results of the work in this thesis will be used to provide some insight into quality issues and criteria. One such example would be international elements of COS development. If the goal of the COS work is to get an international perspective, then of course international representation would be a mark of quality. However, if this is not the goal, or resource limitations mean that COS have to be developed at a more local or national level, then it is currently unclear how this transposes to quality. Further methodological research is needed, to answer some of the questions about methods posed by this research (a research agenda is presented in section 8.6), before we can adequately determine quality. The interviews carried out as part of this thesis also demonstrated that 'a one size fits all' approach to methods is not relevant for COS development, as different populations and disease areas have specific considerations. This makes the assessment of quality particularly challenging. The development of an instrument for quality assessment of COS studies remains a priority on the COMET research agenda.

8.4.3 Funding COS development

Both surveyed and interviewed developers expressed an overarching concern related to resource limitations, particularly the difficulty with obtaining funding for COS development. Interviewed developers were compromised in how they went about COS work as a consequence of limited funding. The acquisition of funding for COS work is therefore critical to avoid compromise in the methodology that could ultimately impact upon the final core set. As such, it is vital to raise awareness about COS and the fact that resources are needed to develop them amongst research funding agencies. Work is ongoing within the COMET Initiative to identify representatives from different countries who are best placed to start such communication with funders in their country. A pilot is underway, with a representative in Portugal, who has started the process of engaging the relevant Portuguese funding agency on COS development research. COMET can then learn from this pilot interaction to build an effective model for engaging research funding agencies. Guidelines should also comment on the important issue of obtaining funding for COS development.

8.4.4 A framework for outcomes categorisation

The extensive systematic review that was undertaken as part of this thesis demonstrated the difficulty with categorising outcomes. The various ways of classifying outcomes and the current uncertainty around which of these is most suitable, coupled with the high number of studies included in the review, meant that it was too complex to categorise outcomes. Nevertheless, classification of outcomes into an agreed framework is highly desirable to help future developers think about potentially relevant outcomes. The existing frameworks for doing this were briefly described in Chapter 1 of this thesis, and it was clear that there is inconsistency between the different models. I therefore recommend a more formal review of the existing frameworks and schemes for categorising outcomes. An appraisal of existing methods would help to identify which, if any, of these existing frameworks and schemes are suitable for categorising outcomes. This would be of benefit to developers when they review the outcomes used in previous research as part of the COS development process.

The COMET Management Group worked with Valerie Smith [29] to compare the review of outcomes from Cochrane reviews, with the outcomes recommended as core in those COS included in the systematic review reported in this thesis. Smith et al's classification was amended based on similar outcomes across different categories and The Cochrane Collaboration is currently trialling that classification in two Cochrane Review Groups, with view to adopting an agreed framework for classification of outcomes.

8.4.5 Implementation and uptake of COS

COS developers who were interviewed as part of this study described implementation of the COS as a major challenge. Implementation of a COS was also presented as a challenge by developers in the survey. The majority (84%) of surveyed developers, when asked whether implementation or uptake was considered during COS development, answered that it was. Those who answered 'no' indicated that they saw implementation as beyond the scope of COS development, or that funding was not available. Both studies also indicated that monitoring the uptake of COS will be challenging.

To my knowledge only two studies have examined the impact of COS, and each have focussed on particular COS. Kirkham et al. [156] carried out an observational review of rheumatoid arthritis trials identified through the Cochrane Library, and evaluated whether

or not there were trends in the proportion of trials reporting the full COS over time. Findings indicated an upward trend in the proportion of trials reporting on the full rheumatoid arthritis COS. The study examined published trial reports but the method proved to be lengthy. Furthermore the study did not include data about recent trials, or indeed that were currently underway. [157]. Bautista-Molano et al. [158] carried out a systematic literature review to investigate how well a COS for ankylosing spondylitis had been implemented in clinical trials according to the type of intervention. Similarly, an increase over time was reported.

COS have the potential to improve the evidence base for health care, but consideration must be given to the methods for disseminating their availability amongst the relevant communities. With the number of COS in development on the increase, it is timely to examine the level of uptake of a wider range of COS to identify an efficient and effective way of monitoring uptake. An MRC HTMR studentship has recently commenced to develop an efficient approach to assessing COS uptake with results based on up to date information [157].

8.5 Methodological guidance for COS development

In addition to the areas already discussed in this chapter, this work has demonstrated an urgent need for methodological guidance and research to inform future COS development. As discussed in Chapter 7, guidance needs to promote awareness of important issues, encouraging developers to think about their own contexts, which will in turn support developers to make methodological decisions that suit their needs and resources. Some of the issues raised might not be unique to COS development, therefore there is a need to look to other disciplines and applications to see what can be learnt from existing research. Methodological guidelines for some aspects of COS methods, such as Delphi or consensus meetings, may well be found in other contexts such as clinical guideline development. This should be explored to avoid unnecessary duplication.

8.5.1 A platform for methodological guidance and research

COS development guidance and methodological research need to be freely available and accessible to developers. The COMET Initiative are developing a handbook which brings together existing research about COS development. The previous issues to consider as

highlighted by Williamson et al [38] will be expanded upon, as well as the inclusion of additional issues identified since their publication. This expansion will be informed by the work in this thesis. The work undertaken here has identified how developers have tackled the challenges of COS development, such as the inclusion of patient participants in the COS development process. This work has also shown that patient participation in COS development has increased, and although it was not originally commonplace, the majority of COS now include patient participation and there is agreement that this is an important component. The question is therefore no longer whether patients should participate, but rather the nature of that involvement (see Table 38 below). Interviewed developers' experiences will be incorporated to provide examples of successful or unsuccessful strategies. The handbook will serve as a resource for developers, and will be updated as new data, evidence and guidance emerges. Because it takes time for new evidence to be published, there is a need for this new information to be made immediately available.

Access to the COMET Initiative website continues to increase and COS awareness is growing internationally [144]. The COMET website is therefore ideally placed to host a dynamic platform for methodological guidance and research. Similar to the existing COMET database that contains published and ongoing COS, a methodological database could be developed to include both published and ongoing material. Furthermore, known gaps in research or methodological uncertainties could also be listed with a contact link to enable individuals or groups to make suggestions or proffer assistance to address unanswered questions. The inclusion of work ongoing and unanswered questions would be particularly important, because as one COS interviewed developer commented, '*answers to methodological questions and uncertainties may not be immediately forthcoming, but raising awareness of important areas and potential pitfalls in itself would be a significant contribution*'.

8.5.2 Research within research

One way to address uncertainties about COS development may be through research within research. SWAT studies, a Study Within a Trial, have been suggested as a way of addressing uncertainty in methods for doing trials [159]. Similarly nested studies within COS development could answer important questions. For example, randomised controlled trials were nested within the development of three core sets (each including a Delphi process with two rounds of questionnaires completed by patients and health professionals) to

examine the impact of receiving feedback from different stakeholders on the subsequent rating of items and level of consensus between stakeholders [160]. Studies within studies to answer research questions will be an efficient and timely way to address research uncertainties. It is unlikely that the COMET Initiative will have the resources to conduct all of the necessary research, but it might be possible for advisors within the COMET team to develop protocols for priority methodology research that developers could then execute. These protocols could also be made available in a methodology database, minimising unnecessary duplication of work and facilitating collaborative efforts. The use of studies nested within COS projects to help resolve uncertainty will generate the evidence needed to support well informed decisions about COS development methods.

8.6 Proposed research agenda and future work

This thesis has identified areas of uncertainty in COS and provided topics for focus for further research to resolve these uncertainties. Taken together the findings from the systematic review, survey and interviews point to the following as questions where further research is required:

1 Stakeholder involvement

1.1 How can COS developers know and demonstrate that the people sampled are 'the right' people?

1.2 How can COS developers incorporate diversity within stakeholder samples?

1.6 Do different stakeholder groups understand the nature of COS and what they are being asked to do, and does this matter?

1.3 How can response rates be improved?

1.4 How can participation be maintained?

1.5 How can COS developers reconcile different stakeholder groups' opinions?

1.6 Is stakeholder input in COS development *participation* in research or *consultation* in research?

2 How can patient participation in COS development be improved?

2.1 Do patient participants understand the nature of COS and what they are being asked to do, and does this matter?

2.2 Should COS developers involve patient advocates, patient participants, or both?

2.3 Is patient input in COS development *participation* in research or *consultation* in research?

2.4 Are there specific challenges of including certain participants such as children and young people as participants in COS development?**

2.5 How can participants such as children and young people be included effectively as participants in COS development?**

3 Systematic review of outcomes

3.1 What studies should be included?

3.2 What are the rules or thresholds for this type of review?

3.3 How should outcomes be categorised?

4 Delphi methodology

4.1 How should consensus be defined?
4.2 How should feedback be given to participants, and how does feedback affect responses?
4.3 Should outcomes be retained or discarded between rounds?

5 Consensus meetings

5.1 How should consensus be defined?
5.2 How should a consensus meeting be conducted?

6 General

6.1 How general or specific should the scope of a COS be?
6.2 What ethics requirements are necessary for COS development?
6.3 Should COS be developed internationally?
6.4 How should questions about outcome prioritisation be framed (i.e. should the question asked be about which outcomes are important or which outcomes are core?)?

As more developers are now including patient participants in COS development, in Table 38 I have explored how uncertainties about patient participation and involvement might be addressed through further research. I have also considered how the COMET Initiative might actively support and facilitate this further research.

Table 38: Methodological uncertainties in patient participation and involvement

	Research question	Research examples	Example of how COMET might help
2	How can patient participation in COS development be improved?	(1) A systematic review of existing patient participation methodology research, not exclusive to COS development, to see what can be learnt from existing research. For example, de Wit has published work on patient participation in the development of PROs in the area of psoriatic arthritis [161]. Here, the authors summarise facilitators for effective involvement of patients in PRO development. (2) Interviews with patient participants who have been involved in COS development to understand their experiences of the process, and find out what they think was done well and what was not done well. This would highlight the distinctive challenges of COS development in relation to patient participation. (3) There is intrinsic value in including the patient in health research, but it is important to evaluate its impact [162]. Case studies could address the impact PPI has in the design, development and dissemination of a COS study.	- List the research gap/question in a methodology database with a contact link to enable individuals or groups to make suggestions or proffer assistance to address unanswered questions. - Develop proposals or protocols to address research questions, with input and advice from the COMET Initiative People and Patient Participation Involvement and Engagement (PoPPIE) Working Group*. - A repository for resources. As an example, the work in this thesis highlighted that recruitment was more successful when researchers had a way of accessing patient participants, such as through patient groups. ISPOR have 'tools for patients' including a list of patient organisations worldwide: http://www.ispor.org/Patients/PatientOrgWorldwide.asp This, and resources like it, should be identified and included as part of COMET resources. Guidance in this area might suggest that patient groups are useful in thinking through strategies to improve patient participation, but should not be considered as the only method of recruitment due to issues around self-selection of members and a potential lack of representativeness.
2.1	Are there specific challenges of including children and young people as participants in COS development?**	(1) Interviews with COS developers about these specific issues. (2) Qualitative work with children and young people about their experience of being involved in COS work, or their opinion about whether they think they would be able to participate in COS development. (3) A systematic review of broader literature to explore whether work has been done to develop consensus with children and young people in other areas, for example in social sciences or education. What lessons can be learned from other areas?	- Identify published or prospective COS studies that have involved children and young people, and facilitate communication between COS developers and researchers.
2.2	How can children and young people be included effectively as participants in COS development?**		

2.3	Do patient participants understand the nature of COS and what they are being asked to do, and does this matter?***	(1) Qualitative research to explore how people talk about outcomes – we need to understand the language that people use to talk about outcomes as they do not tend to use the word 'outcome'. This could potentially be addressed by exploring how people discuss outcomes in social media. (2) Jepson et al found that when trial recruiters explained randomisation to patients, patients wanted to know 'why' they were being randomised above all else [163]. A survey of COS developers could be conducted, about how COS and instructions are given to patient participants, to identify how a COS as a concept is explained. A survey could also be carried out with corresponding patient participants to see how the concept of a COS was understood. This would help to highlight whether the focus is on patients' understanding or developers' ability to explain clearly and meaningfully to patients. (3) A randomised study within COS development to explore different ways of explaining COS to patient participants and the impact this has on patient participants' understanding and participation.	- Develop guidance about how COS should be described for patients. This might include materials that will help to facilitate researchers' discussions of COS. The use of audio-visual aids (e.g. a video) for recruitment would mean that basic information is delivered consistently every time. Personalisation of information should also be considered and ways of responding to individualised questions and needs for information.
2.4	Should COS developers include patient advocates, patient participants, or both?	(1) Review the literature in more detail to identify how patient participation has been conducted to date, particularly in ongoing COS development. The ethical approval requirements for COS development should also be considered. (2) A Delphi process including COS developers and COS patient participants could be undertaken to agree an optimum model for involving the patient in COS development. This might include the stage of development as well as the number, or proportion, of participants that should be included at each stage.	- Draw on the experience of PoPPIE group members of patient participation and involvement in research to provide guidance.
2.5	Is patient participation in COS development considered to be participation in research or consultation in research?		

** PoPPIE is a Working Group set up to lead and oversee the patient participation, involvement and engagement work of the COMET Initiative. They held their inaugural meeting in November 2015 to discuss the importance of patient and public involvement in developing and overseeing core outcome set studies and the wider activities of patient and public engagement in relation to the work of the COMET Initiative.*

*** As discussed in Chapter 6, whilst there might be important or distinctive methodological considerations for adults with additional needs (e.g. older adults, people with learning disabilities) as well as for children, the number of studies in the review was too few to make this a specific focus of the work in this thesis. Therefore, methodological considerations for other groups is a necessary area of exploration for further research.*

**** The interviews with developers highlighted that problems with understanding outcomes and COS was not exclusive to patients. Therefore, the issue of understanding should be explored with other stakeholder groups as well as patients.*

8.7 Improving the search for COS

The COS identified in the systematic review chapter have formed the basis of an online searchable database. A short pop-up survey in 2015, to ascertain why people were searching the COMET database, showed that the most common reasons for searching the database were to inform decision making about developing a COS, or to inform the outcomes in planning a clinical trial [164]. The database was also searched by systematic reviewers, funders of COS and trials, and people who had been asked to take part in the development of a COS study. The results of the pop up survey, in particular that people thinking about developing a COS are checking the COMET database to see whether a COS exists in their area of interest to avoid duplication, emphasise the importance of keeping the database current. To achieve this, the search for COS needs to be performed on a regular basis.

The challenges of performing this search have been discussed extensively in this thesis. One way of addressing these challenges would be the introduction of a COS MeSH term in MEDLINE and other bibliographic databases for identifying COS papers. However, the inclusion of a new MeSH term in itself might be a challenge. MeSH is the National Library of Medicine's controlled vocabulary thesaurus, who state that before a new descriptor is introduced, consideration must be given to how much is published about that topic; if little is published the Library see little purpose or advantage in creating a new descriptor in a vocabulary which has to encompass the subject content of the entire published literature [165]. Although COS development is a growing area of research, it is relatively niche when compared to the broad nature of most medical subject headings. It is therefore probable that the proposal of a COS heading is unrealistic at this time.

In the absence of a COS MeSH term, standards for reporting, as described in section 8.4.1, could help with the searching and identification of COS development studies. The use of standardised terms, for example 'core outcome set' in the title and abstract would make it much easier to identify appropriate COS development papers. Although under development, it will take some time for the production, publication and adoption of the reporting guideline for COS development studies to benefit the searching of these papers. In the interim, I therefore recommend a review of the current search strategy to see where improvements can be made immediately. The systematic review of COS has already been updated once [164], and a second update is underway, both using the original search. As

with the original search, the update proved to be resource intensive due to the large retrieval of records that need assessing. I recommend a review of the language used in the COS papers identified in the updated searches, to see if there is already a trend towards using more consistent language to describe COS development. A comparison could then be performed with the search terms to see where improvements could be made. If accuracy could be retained whilst significantly reducing the number of hits, and therefore the amount of irrelevant material needed to read, this would be advantageous to maintaining an up to date database of COS.

8.8 Wider implications

The 'E' in COMET stands for effectiveness, and that has been the focus of the work undertaken in this thesis. However, COS should not be completely disparate to clinical practice. Outcomes that are important in effectiveness trials and research might be important for routine clinical practice and vice versa. The potential for using electronic health records in health research have been well documented [166], and there is a growing interest in e-trials using health records to collect outcome data. The inclusion of core outcomes in routine health records should therefore be considered to maximise efficiency in e-trials. Although presently very few trials run entirely based in routine health records, this is a vision for the future. As such, consideration of including core outcomes in routine health records now will help towards achieving this vision whilst avoiding future complications. There is also a need to critique e-trials where a COS exists, but that have been run from the health record alone and may not have included the most important outcomes. E-trial designers need to think carefully about the outcomes included in routine collection. For all trials, researchers should start by looking at the COMET database to see whether a COS exists. If there is a COS, a cross-check could be performed to see whether the routine health record collects data on those outcomes. If it does not, then researchers should consider how best to collect the missing information. There is an overall need to collaborate on health research and clinical practice infrastructure, working towards agreement about what is important to measure, collect and report routinely.

The importance of COS methodology can also be seen from its potential application to other areas. For example, similar methodology has recently been applied to develop core information sets for the disclosure of information in consultations with patients before surgery [167]. The use of core information sets to help patients understand what to expect

from surgery demonstrates the significance of the methodology used to arrive at this core set. Methodology used to develop core information sets might be improved as direct result of the work described in this thesis. Additionally, future work to develop core information sets might also provide additional opportunities to answer methodological questions of relevance to COS development.

The value that is placed on COS is of utmost importance to their success. There was evidence of some interviewees attaching little value to their COS development work, as well as struggling to gain acceptance of their work within their wider communities. The work undertaken in this thesis has suggested that value is linked to understanding about COS, as well as the unease that comes with working with opinion based data. Whist guidance and further research around methods will go some way to help increase the perceived value of COS; guidance alone will not foster value. To enhance the value attributed to COS development we might be able to learn from analogous examples, such as the increased emphasis on PPI in health research in the last decade. PPI has increased in perceived value, and it would be advantageous to consider how this has been achieved and if there are any lessons for COS development. One explanation is the backing that PPI received from important stakeholder groups, organisations and influential individuals. For example, a report commissioned by Cancer Research UK [168], evaluating the current state of research in the NHS, described the NIHR as having a central and enabling role in the increased involvement of patients in research. This demonstrates the importance of the role of research funders, such as NIHR, in the transformation of research practices. Furthermore, the support of individuals such as the Chief Medical Officer (Dame Sally Davies) has also given value to PPI in health research [169]. Funders of research are starting to acknowledge the importance of COS, with the NIHR HTA programme in the UK, the Health Research Board in Ireland and the charity Arthritis Research UK, now encouraging applicants to consider COS when seeking funding for new trials. As described in section 8.4.3, COMET is undertaking work to advance engagement with funders. Furthermore, groups like CROWN now have over 70 journals involved that are committed to promoting the uptake of COS in the area of women's health [170], are likely to add value to COS in that disease area. As CROWN is a new initiative, the impact will need to be evaluated in time; but, if successful, this might be a model that other areas could replicate. Understanding how the perceived value of COS can be increased will facilitate the development and application of COS in health research.

8.9 Final summary

To make well informed decisions about healthcare, we need to be able to compare and contrast research findings on the basis of the same outcomes. COS represent the minimum important outcomes that should be measured and reported in all trials for a specific condition. For COS to be successfully implemented, they need to be easily accessible to researchers and other key groups, developed using rigorous methods, and reported clearly. The work in this thesis has contributed to that goal by bringing COS together in one place for the first time. These results provide the first comprehensive account of COS development, and will inform the formulation of much needed guidance in this area, thus improving COS development methodology. The use of COS in research means that all trials in a given area of health are measuring and reporting the agreed important outcomes. This will improve the quality of evidence used in health care decision making, ultimately translating to improved health care for patients. It is therefore critical that COS are developed in the best possible way. The answers to the methodological questions and uncertainties posed by this research may not be immediately forthcoming, but by raising awareness of important areas and potential pitfalls, the work undertaken in this thesis will make a valuable contribution to COS knowledge and future research.

List of references

1. Gartlehner, G., R.A. Hansen, D. Nissman, K.N. Lohr, et al., *Criteria for Distinguishing Effectiveness From Efficacy Trials in Systematic Reviews*2006, Rockville MD.
2. Akobeng, A.K., *Understanding randomised controlled trials.* Arch Dis Child, 2005. **90**(8): p. 840-844.
3. FDA. *Clinical Studies Section of Labeling for Human Prescription Drug and Biological Products — Content and Format: Guidance for Industry* 2006 [cited 2016 May]; Available from: http://www.fda.gov/downloads/RegulatoryInformation/Guidances/ucm127534.pdf.
4. NICE. *The guidelines manual* 2012 [cited 2016 May]; Available from: https://www.nice.org.uk/article/pmg6/chapter/1-Introduction.
5. Flay, B.R., *Efficacy and effectiveness trials (and other phases of research) in the development of health promotion programs.* Preventive Medicine, 1986. **15**(5): p. 451-474.
6. Godwin, M., L. Ruhland, I. Casson, S. MacDonald, et al., *Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity.* BMC Med Res Methodol, 2003. **3**: p. 28.
7. Meinert, C.L., *Clinical Trials - Design, Conduct and Analysis.* Second edition ed2012, New York: Oxford University Press.
8. Sackett, D., W. Richardson, W. Rosenberg, and R. Haynes, *Evidence-based medicine: How to practice and teach EBM*, ed. C. Livingston1997, New York.
9. *ICH Harmonised Tripartite Guideline. Statistical principles for clinical trials. International Conference on Harmonisation E9 Expert Working Group.* Stat Med, 1999. **18**(15): p. 1905-42.
10. Sedgwick, P., *Primary and secondary outcome measures.* BMJ, 2010. **340**.
11. Moher, D., S. Hopewell, K.F. Schulz, V. Montori, et al., *CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials.* BMJ, 2010. **340**: p. c869.
12. Higgins, J. and S. Green, *[Editors]. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [Updated March 2011].* 2011.
13. AHRQ, *Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide*, ed. P. Velentgas, et al.2013, Rockville MD: Agency for Healthcare Research and Quality.
14. Grimes, D.A. and K.F. Schulz, *Surrogate end points in clinical research: hazardous to your health.* Obstet Gynecol, 2005. **105**(5 Pt 1): p. 1114-8.
15. Aronson, J.K., *Biomarkers and surrogate endpoints.* Br J Clin Pharmacol, 2005. **59**(5): p. 491-4.
16. Strimbu, K. and J.A. Tavel, *What are biomarkers?* Curr Opin HIV AIDS, 2010. **5**(6): p. 463-6.
17. Zarin, D.A., T. Tse, R.J. Williams, R.M. Califf, et al., *The ClinicalTrials.gov results database--update and key issues.* N Engl J Med, 2011. **364**(9): p. 852-60.
18. Cordoba, G., L. Schwartz, S. Woloshin, H. Bae, et al., *Definition, reporting, and interpretation of composite outcomes in clinical trials: systematic review.* BMJ, 2010. **341**.
19. Bakas, T., S.M. McLennon, J.S. Carpenter, J.M. Buelow, et al., *Systematic review of health-related quality of life models.* Health Qual Life Outcomes, 2012. **10**: p. 134.

20. WHO. *WHO definition of Health. Geneva 1948.* [cited 2015 28th April]; Available from: http://www.who.int/about/definition/en/print.html.
21. Idzerda, L., T. Rader, P. Tugwell, and M. Boers, *Can We Decide Which Outcomes Should Be Measured in Every Clinical Trial? A Scoping Review of the Existing Conceptual Frameworks and Processes to Develop Core Outcome Sets.* J Rheumatol, 2014.
22. NIH. *Patient-Reported Outcomes Measurement Information System - PROMIS.* [cited 2015 28th April]; Available from: http://www.nihpromis.org/measures/domainframework.
23. Kostanjsek, N., *Use of The International Classification of Functioning, Disability and Health (ICF) as a conceptual framework and common language for disability statistics and health information systems.* BMC Public Health, 2011. **11 Suppl 4**: p. S3.
24. Fries, J.F., P. Spitz, R.G. Kraines, and H.R. Holman, *Measurement of patient outcome in arthritis.* Arthritis & Rheumatism, 1980. **23**(2): p. 137-45.
25. Wilson, I.B. and P.D. Cleary, *Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes.* JAMA, 1995. **273**(1): p. 59-65.
26. Ferrans, C.E., J.J. Zerwic, J.E. Wilbur, and J.L. Larson, *Conceptual model of health-related quality of life.* J Nurs Scholarsh, 2005. **37**(4): p. 336-42.
27. Boers, M., J.R. Kirwan, G. Wells, D. Beaton, et al., *Developing core outcome measurement sets for clinical trials: OMERACT filter 2.0.* Journal of Clinical Epidemiology, 2014. **67**(7): p. 745-53.
28. Gliklich, R.E., M.B. Leavy, J. Karl, D.M. Campion, et al., *A framework for creating standardized outcome measures for patient registries.* J Comp Eff Res, 2014. **3**(5): p. 473-80.
29. Smith, V., M. Clarke, P. Williamson, and E. Gargon, *Survey of new 2007 and 2011 Cochrane reviews found 37% of prespecified outcomes not reported.* Journal of Clinical Epidemiology, 2014.
30. Chalmers, I. and P. Glasziou, *Avoidable waste in the production and reporting of research evidence.* Lancet, 2009. **374**(9683): p. 86-9.
31. Hirsch, B.R., R.M. Califf, S.K. Cheng, A. Tasneem, et al., *Characteristics of oncology clinical trials: insights from a systematic analysis of ClinicalTrials.gov.* JAMA Intern Med, 2013. **173**(11): p. 972-9.
32. Kirkham, J.J., E. Gargon, M. Clarke, and P.R. Williamson, *Can a core outcome set improve the quality of systematic reviews?--a survey of the Co-ordinating Editors of Cochrane Review Groups.* Trials, 2013. **14**: p. 21.
33. Miyar, J. and C.E. Adams, *Content and quality of 10,000 controlled trials in schizophrenia over 60 years.* Schizophr Bull, 2013. **39**(1): p. 226-9.
34. Hutton, J.L. and P.R. Williamson, *Bias in meta-analysis due to outcome variable selection within studies.* Applied Statistics, 2000. **49**: p. 359-70.
35. Dwan, K., D.G. Altman, J.A. Arnaiz, J. Bloom, et al., *Systematic review of the empirical evidence of study publication bias and outcome reporting bias.* PLoS ONE [Electronic Resource], 2008. **3**(8): p. e3081.
36. Kirkham, J.J., K.M. Dwan, D.G. Altman, C. Gamble, et al., *The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews.* BMJ, 2010. **340**: p. c365.
37. Chan, A.W., F. Song, A. Vickers, T. Jefferson, et al., *Increasing value and reducing waste: addressing inaccessible research.* Lancet, 2014. **383**(9913): p. 257-66.
38. Williamson, P.R., D.G. Altman, J.M. Blazeby, M. Clarke, et al., *Developing core outcome sets for clinical trials: issues to consider.* Trials, 2012. **13**(1): p. 132.

39. Clarke, M., *Standardising outcomes for clinical trials and systematic reviews.* Trials, 2007. **8**: p. 39.
40. Gargon, E., P.R. Williamson, D.G. Altman, J.M. Blazeby, et al., *The COMET Initiative database: progress and activities from 2011 to 2013.* Trials, 2014. **15**(1): p. 279.
41. Miller, A.B., B. Hoogstraten, M. Staquet, and A. Winkler, *Reporting results of cancer treatment.* Cancer, 1981. **47**(1): p. 207-14.
42. Boers, M., P. Tugwell, D.T. Felson, P.L. van Riel, et al., *World Health Organization and International League of Associations for Rheumatology core endpoints for symptom modifying antirheumatic drugs in rheumatoid arthritis clinical trials.* Journal of Rheumatology - Supplement, 1994. **41**: p. 86-9.
43. Tugwell, P., M. Boers, P. Brooks, L. Simon, et al., *OMERACT: an international initiative to improve outcome measurement in rheumatology.* Trials, 2007. **8**: p. 38.
44. Kirwan, J.R., S.E. Hewlett, T. Heiberg, R.A. Hughes, et al. *Incorporating the patient perspective into outcome assessment in rheumatoid arthritis - progress at OMERACT 7*. Journal of Rheumatology, 2005. 2250-6.
45. Kirwan, J., T. Heiberg, S. Hewlett, R. Hughes, et al., *Outcomes from the Patient Perspective Workshop at OMERACT 6.* Journal of Rheumatology, 2003. **30**(4): p. 868-72.
46. Kirwan, J.R., P. Minnock, A. Adebajo, B. Bresnihan, et al., *Patient perspective: fatigue as a recommended patient centered outcome measure in rheumatoid arthritis.* Journal of Rheumatology, 2007. **34**(5): p. 1174-7.
47. Khan, K., *The CROWN Initiative: journal editors invite researchers to develop core outcomes in women's health.* J Ovarian Res, 2015. **8**(1): p. 6.
48. ICHOM. *Our standard sets*. [cited 2015 28th April]; Available from: http://www.ichom.org/medical-conditions.
49. Martin, N.E., L. Massey, C. Stowell, C. Bangma, et al., *Defining a standard set of patient-centered outcomes for men with localized prostate cancer.* Eur Urol, 2015. **67**(3): p. 460-7.
50. MacLennan, S., P.R. Williamson, and T.B. Lam, *Re: Neil E. Martin, Laura Massey, Caleb Stowell, et al. Defining a Standard Set of Patient-centered Outcomes for Men with Localized Prostate Cancer. Eur Urol 2015;67:460–7.* Eur Urol, 2015.
51. Mokkink, L.B., C.B. Terwee, D.L. Patrick, J. Alonso, et al., *The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study.* Quality of Life Research, 2010. **19**(4): p. 539-49.
52. Prinsen, C.A., S. Vohra, M.R. Rose, S. King-Jones, et al., *Core Outcome Measures in Effectiveness Trials (COMET) initiative: protocol for an international Delphi study to achieve consensus on how to select outcome measurement instruments for outcomes included in a 'core outcome set'.* Trials, 2014. **15**(1): p. 247.
53. Ioannidis, J.P.A., S. Greenland, M.A. Hlatky, M.J. Khoury, et al., *Increasing value and reducing waste in research design, conduct, and analysis.* The Lancet, 2014. **383**(9912): p. 166-175.
54. Simera, I. and D.G. Altman, *Reporting medical research.* Int J Clin Pract, 2013. **67**(8): p. 710-6.
55. Schulz, K.F., D.G. Altman, D. Moher, and C. Group, *CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials.* PLoS Med, 2010. **7**(3): p. e1000251.
56. Chan, A.W., J.M. Tetzlaff, P.C. Gotzsche, D.G. Altman, et al., *SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials.* BMJ, 2013. **346**: p. e7586.

57. Glasziou, P., D.G. Altman, P. Bossuyt, I. Boutron, et al., *Reducing waste from incomplete or unusable reports of biomedical research.* The Lancet, 2014. **383**(9913): p. 267-276.
58. Sinha, I., L. Jones, R.L. Smyth, and P.R. Williamson, *A systematic review of studies that aim to determine which outcomes to measure in clinical trials in children.* PLoS Med, 2008. **5**(4): p. e96.
59. Sinha, I., R.L. Smyth, and P.R. Williamson, *Using the delphi technique to determine which outcomes to measure in clinical trials: recommendations for the future based on a systematic review of existing studies.* PLoS Med, 2011. **8**(1): p. e1000393.
60. O'Cathain, A., *Reporting mixed methods projects*, in *Mixed methods research for nursing and the health sciences*, S. Andrew and E.J. Halcomb, Editors. 2009, Blackwell Publishing Ltd United Kingdom.
61. Oliver, S. and J. Gray, *A bibliography of research reports about patients', clinicians' and researchers' priorities for new research.* London: James Lind Alliance. **December 2006**.
62. *Biochemical verification of tobacco use and cessation.* Nicotine Tob Res, 2002. **4**(2): p. 149-59.
63. Schlaepfer, T.E. and J.J. Fins, *Deep brain stimulation and the neuroethics of responsible publishing: when one is not enough.* JAMA, 2010. **303**(8): p. 775-6.
64. McAllister, M., L. Davies, K. Payne, S. Nicholls, et al., *The emotional effects of genetic diseases: implications for clinical genetics.* American Journal of Medical Genetics Part A, 2007. **143A**(22): p. 2651-61.
65. McAllister, M., G. Dunn, and C. Todd, *Empowerment: qualitative underpinning of a new clinical genetics-specific patient-reported outcome.* European Journal of Human Genetics, 2011. **19**(2): p. 125-30.
66. McAllister, M., K. Payne, R. Macleod, S. Nicholls, et al., *Patient empowerment in clinical genetics services.* Journal of Health Psychology, 2008. **13**(7): p. 895-905.
67. McAllister, M., K. Payne, S. Nicholls, R. MacLeod, et al., *Improving service evaluation in clinical genetics: identifying effects of genetic diseases on individuals and families.* Journal of Genetic Counseling, 2007. **16**(1): p. 71-83.
68. McAllister, M., A.M. Wood, G. Dunn, S. Shiloh, et al., *The Genetic Counseling Outcome Scale: a new patient-reported outcome measure for clinical genetics services.* Clinical Genetics, 2011. **79**(5): p. 413-24.
69. Payne, K., S. Nicholls, M. McAllister, R. Macleod, et al., *Outcome measurement in clinical genetics services: a systematic review of validated measures.* Value Health, 2008. **11**(3): p. 497-508.
70. Hill, B., *Comparison of journal title coverage between CINAHL and Scopus.* J Med Libr Assoc, 2009. **97**(4): p. 313-4.
71. Watson, R.J. and P.H. Richardson, *Identifying randomized controlled trials of cognitive therapy for depression: comparing the efficiency of Embase, Medline and PsycINFO bibliographic databases.* Br J Med Psychol, 1999. **72 (Pt 4)**: p. 535-42.
72. Wong, S.S., N.L. Wilczynski, and R.B. Haynes, *Comparison of top-performing search strategies for detecting clinically sound treatment studies and systematic reviews in MEDLINE and EMBASE.* J Med Libr Assoc, 2006. **94**(4): p. 451-5.
73. Hopewell, S., J. Clarke Mike, C. Lefebvre, and W. Scherer Roberta *Handsearching versus electronic searching to identify reports of randomized trials*. Cochrane Database of Systematic Reviews, 2007. DOI: 10.1002/14651858.MR000001.pub2.
74. Higgins JPT and S. Green, *Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.2 [updated September 2009].* The Cochrane Collaboration, 2009.

75. Clarke, M., *Searching MEDLINE for randomised trials.* BMJ, 1993. **307**(6903): p. 565-565.
76. Dickersin, K., E. Manheimer, S. Wieland, K.A. Robinson, et al., *Development of the Cochrane Collaboration's CENTRAL Register of controlled clinical trials.* Eval Health Prof, 2002. **25**(1): p. 38-64.
77. Lefebvre, C., A. Eisinga, S. McDonald, and N. Paul, *Enhancing access to reports of randomized trials published world-wide--the contribution of EMBASE records to the Cochrane Central Register of Controlled Trials (CENTRAL) in The Cochrane Library.* Emerg Themes Epidemiol, 2008. **5**: p. 13.
78. INVOLVE, *Briefing notes for researchers: involving the public in NHS, public health and social care research*, INVOLVE, Editor 2012: Eastleigh.
79. Moher, D., A. Liberati, J. Tetzlaff, D.G. Altman, et al., *Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement.* PLoS Med, 2009. **6**(7): p. e1000097.
80. Giacoia, G.P., D.L. Birenbaum, H.C. Sachs, and D.R. Mattison, *The newborn drug development initiative.* Pediatrics, 2006. **117**(3 Pt 2): p. S1-8.
81. Cranney, A., P. Tugwell, S. Cummings, P. Sambrook, et al., *Osteoporosis clinical trials endpoints: candidate variables and clinimetric properties.* Journal of Rheumatology, 1997. **24**(6): p. 1222-9.
82. Moher, D., K.F. Schulz, I. Simera, and D.G. Altman, *Guidance for developers of health research reporting guidelines.* PLoS Med, 2010. **7**(2): p. e1000217.
83. Murphy, M.K., N.A. Black, D.L. Lamping, C.M. McKee, et al., *Consensus development methods, and their use in clinical guideline development.* Health Technol Assess, 1998. **2**(3): p. i-iv, 1-88.
84. Bowling, A., *Research Methods in Health.*2009, London: Open University Press.
85. Kitzinger, J., *Qualitative Research: Introducing focus groups.* BMJ, 1995. **311**(7000): p. 299-302.
86. Perry, S. and J.T. Kalberer, Jr., *The NIH consensus-development program and the assessment of health-care technologies: the first two years.* N Engl J Med, 1980. **303**(3): p. 169-72.
87. Stocking, B., Jennett, B. Spiby, J. , *Criteria for change. The history and impact of consensus development conferences in the UK.*1991, London: King's Fund Centre.
88. McGlynn, E.A., J. Kosecoff, and R.H. Brook, *Format and conduct of consensus development conferences. Multi-nation comparison.* Int J Technol Assess Health Care, 1990. **6**(3): p. 450-69.
89. Ackerman, A.B., *A critique of an N.I.H. Consensus Development Conference about "early" melanoma.* Am J Dermatopathol, 1993. **15**(1): p. 52-8.
90. Dalkey, N.C., *The Delphi method: An experimental study of group opinion.* The RAND Corp, 1969. **RM-5888-PR**.
91. Sackman, H., *Delphi Critique*. Vol. Lexington, MA. 1975: Lexington Books.
92. Pill, J., *The Delphi method: Substance, context, a critique and an annotated bibliography.* Socio-Economic Planning Sciences, 1971. **5**(1): p. 57-71.
93. Goodman, C.M., *The Delphi technique: a critique.* Journal of Advanced Nursing, 1987. **12**(6): p. 729-34.
94. Hasson, F., S. Keeney, and H. McKenna, *Research guidelines for the Delphi survey technique.* Journal of Advanced Nursing, 2000. **32**(4): p. 1008-15.
95. Boynton, P.M. and T. Greenhalgh, *Selecting, designing, and developing your questionnaire.* BMJ, 2004. **328**(7451): p. 1312-1315.
96. Delbecq, A.L., Van de Ven, A.H., *A Group Process Model for Problem Identification and Program Planning.* Journal of Applied Behavioral Science, 1971. **7**: p. 466.

97. Potter, M., Gordon, S., Hamer, P. , *The Nominal Group Technique: A useful consensus methodology in physiotherapy research.* NZ Journal of Physiotherapy, 2004. **32**(3): p. 126-130.
98. Chasens, E.R. and E. Olshansky, *Daytime sleepiness, diabetes, and psychological well-being.* Issues in Mental Health Nursing, 2008. **29**(10): p. 1134-50.
99. Harvey, N. and C.A. Holmes, *Nominal group technique: An effective method for obtaining group consensus.* Int J Nurs Pract, 2012. **18**(2): p. 188-94.
100. Green, J., Thorogood,N. , *Qualitative Methods for Health Research.*2008, London: SAGE Publications Ltd.
101. Uman, L.S., *Systematic Reviews and Meta-Analyses.* Journal of the Canadian Academy of Child and Adolescent Psychiatry, 2011. **20**(1): p. 57-59.
102. Roberts, L. and C. Counsell, *Assessment of clinical outcomes in acute stroke trials.* Stroke, 1998. **29**(5): p. 986-91.
103. Sinha, I.P., R. Gallagher, P.R. Williamson, and R.L. Smyth, *Development of a core outcome set for clinical trials in childhood asthma: a survey of clinicians, parents, and young people.* Trials, 2012. **13**: p. 103.
104. Reddel, H.K., D.R. Taylor, E.D. Bateman, L.P. Boulet, et al., *An official American Thoracic Society/European Respiratory Society statement: asthma control and exacerbations: standardizing endpoints for clinical asthma trials and clinical practice.* Am J Respir Crit Care Med, 2009. **180**(1): p. 59-99.
105. Busse, W.W., W.J. Morgan, V. Taggart, and A. Togias, *Asthma outcomes workshop: overview.* Journal of Allergy & Clinical Immunology, 2012. **129**(3 Suppl): p. S1-8.
106. Johnson, R.B., A.J. Onwuegbuzie, and L.A. Turner, *Toward a Definition of Mixed Methods Research.* Journal of Mixed Methods Research, 2007. **1**(2): p. 112-133.
107. Bryman, A., *Social research methods*. Fourth edition ed2012, New York: Oxford University Press Inc.
108. Creswell, J.W. and V. Plano Clark, *Designing and Conducting Mixed Methods Research*2007: SAGE Publications, Inc.
109. Edwards, P.J., I. Roberts, M.J. Clarke, C. Diguiseppi, et al., *Methods to increase response to postal and electronic questionnaires.* Cochrane Database Syst Rev, 2009(3): p. MR000008.
110. Evans, J.R. and M. Anil, *The value of online surveys.* Internet Research, 2005. **15**(2): p. 195-219.
111. Fricker, R.D. and M. Schonlau, *Advantages and Disadvantages of Internet Research Surveys: Evidence from the Literature.* Field Methods, 2002. **14**(4): p. 347-367.
112. Dykema, J., N.R. Jones, T. Piche, and J. Stevenson, *Surveying clinicians by web: current issues in design and administration.* Eval Health Prof, 2013. **36**(3): p. 352-81.
113. Popay, J. and G. Williams, *Qualitative research and evidence-based healthcare.* J R Soc Med, 1998. **91 Suppl 35**: p. 32-7.
114. Quinn Patton, M., *Qualitative Research and Evaluation Methods: Integrating Theory and Practice*. Fourth Edition ed2015, United States of America: SAGE Publications, Inc.
115. Ritchie, J., J. Lewis, C. McNaughton Nicholls, and R. Ormston, *Qualitative research practice: A guide for social science students and researchers.*2014, London: Sage.
116. Glaser, B.G., *The Constant Comparative Method of Qualitative Analysis.* Social Problems, 1965. **12**(4): p. 436-445.
117. Thorne, S., *Data analysis in qualitative research.* Evidence Based Nursing, 2000. **3**(3): p. 68-70.
118. Sturges, J.E. and K.J. Hanrahan, *Comparing Telephone and Face-to-Face Qualitative Interviewing: a Research Note.* Qualitative Research, 2004. **4**(1): p. 107-118.

119. McColl, E., A. Jacoby, L. Thomas, J. Soutter, et al., *Design and use of questionnaires: a review of best practice applicable to surveys of health service staff and patients.* Health Technol Assess, 2001. **5**(31): p. 1-256.
120. Baker, S.E. and R. Edwards, *How many qualitative interviews is enough. Discussion Paper.* National Centre for Research Methods. , 2012. **http://eprints.ncrm.ac.uk/2273/**.
121. Kirwan, J.R., S.E. Hewlett, T. Heiberg, R.A. Hughes, et al., *Incorporating the patient perspective into outcome assessment in rheumatoid arthritis--progress at OMERACT 7.* J Rheumatol, 2005. **32**(11): p. 2250-6.
122. Caldwell, P.H., S.B. Murphy, P.N. Butow, and J.C. Craig, *Clinical trials in children.* Lancet, 2004. **364**(9436): p. 803-11.
123. Smyth, R., L., *Research with children.* BMJ, 2001. **322**(7299): p. 1377-1378.
124. McGrath, P.J., G.A. Walco, D.C. Turk, R.H. Dworkin, et al., *Core outcome domains and measures for pediatric acute and chronic/recurrent pain clinical trials: PedIMMPACT recommendations.* Journal of Pain, 2008. **9**(9): p. 771-83.
125. Turk, D.C., R.H. Dworkin, R.R. Allen, N. Bellamy, et al., *Core outcome domains for chronic pain clinical trials: IMMPACT recommendations.* Pain, 2003. **106**(3): p. 337-45.
126. Macefield, R.C., C.E. Boulind, and J.M. Blazeby, *Selecting and measuring optimal outcomes for randomised controlled trials in surgery.* Langenbecks Arch Surg, 2013.
127. Montgomery, J.H., M. Byerly, T. Carmody, B. Li, et al., *An analysis of the effect of funding source in randomized clinical trials of second generation antipsychotics for the treatment of schizophrenia.* Control Clin Trials, 2004. **25**(6): p. 598-612.
128. Seers, K., *Qualitative data analysis.* Evidence Based Nursing, 2012. **15**(1): p. 2.
129. Nicholas, M. and P. Catherine, *Assessing quality in qualitative research.* BMJ, 2000. **320**(7226): p. 50-52.
130. Gale, N.K., G. Heath, E. Cameron, S. Rashid, et al., *Using the framework method for the analysis of qualitative data in multi-disciplinary health research.* BMC Med Res Methodol, 2013. **13**: p. 117.
131. Smith, J. and J. Firth, *Qualitative data analysis: the framework approach.* Nurse Res, 2011. **18**(2): p. 52-62.
132. QSRInternational. *NVivo*. [cited 2016 May]; Available from: http://www.qsrinternational.com/.
133. Barbour, R.S., *Checklists for improving rigour in qualitative research: a case of the tail wagging the dog?* BMJ, 2001. **322**(7294): p. 1115-7.
134. Williams, B., *Checklists for improving rigour in qualitative research. Including personal reflections might help.* BMJ, 2001. **323**(7311): p. 515.
135. Mays, N. and C. Pope, *Qualitative research in health care. Assessing quality in qualitative research.* BMJ, 2000. **320**(7226): p. 50-2.
136. Salmon, P., S. Peters, A. Rogers, L. Gask, et al., *Peering through the barriers in GPs' explanations for declining to participate in research: the role of professional autonomy and the economy of time.* Fam Pract, 2007. **24**(3): p. 269-75.
137. Stige, B., K. Malterud, and T. Midtgarden, *Toward an agenda for evaluation of qualitative research.* Qual Health Res, 2009. **19**(10): p. 1504-16.
138. O'Brien, B.C., I.B. Harris, T.J. Beckman, D.A. Reed, et al., *Standards for reporting qualitative research: a synthesis of recommendations.* Acad Med, 2014. **89**(9): p. 1245-51.
139. Tong, A., P. Sainsbury, and J. Craig, *Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups.* Int J Qual Health Care, 2007. **19**(6): p. 349-57.

140. Guyatt, G.H., A.D. Oxman, G.E. Vist, R. Kunz, et al., *GRADE: an emerging consensus on rating quality of evidence and strength of recommendations.* BMJ, 2008. **336**(7650): p. 924-6.
141. Guyatt, G.H., A.D. Oxman, R. Kunz, Y. Falck-Ytter, et al., *Going from evidence to recommendations.* BMJ, 2008. **336**(7652): p. 1049-51.
142. Maarten Boers, J.R. Kirwan, P. Tugwell, D. Beaton, et al. *The OMERACT Handbook*. 2015; Available from: http://www.omeract.org/pdf/OMERACT_Handbook.pdf.
143. Schmitt, J., C. Apfelbacher, P.I. Spuls, K.S. Thomas, et al., *The Harmonizing Outcome Measures for Eczema (HOME) roadmap: a methodological framework to develop core sets of outcome measurements in dermatology.* Journal of Investigative Dermatology, 2015. **135**(1): p. 24-30.
144. Gargon, E., P.R. Williamson, D.G. Altman, J.M. Blazeby, et al., *The COMET initiative database: progress and activities update (2014).* Trials, 2015. **16**(1): p. 515.
145. Greenwood-Lee, J., P. Hawe, A. Nettel-Aguirre, A. Shiell, et al., *Complex intervention modelling should capture the dynamics of adaptation.* BMC Med Res Methodol, 2016. **16**(1): p. 51.
146. Plsek, P.E. and T. Greenhalgh, *Complexity science: The challenge of complexity in health care.* BMJ, 2001. **323**(7313): p. 625-8.
147. Holland, J.H., *Hidden order: how adaptation builds complexity*1995, Reading, Massachusetts: Perseus Books.
148. Gibson, A., K. Boddy, K. Maguire, and N. Britten, *Exploring the impact of providing evidence-based medicine training to service users.* Research Involvement and Engagement, 2015. **1**(1): p. 1-13.
149. Nijsten, T., T. Rolstad, S.R. Feldman, and R.S. Stern, *Members of the national psoriasis foundation: more extensive disease and better informed about treatment options.* Archives of Dermatology, 2005. **141**(1): p. 19-26.
150. Ives, J., S. Damery, and S. Redwod, *PPI, paradoxes and Plato: who's sailing the ship?* J Med Ethics, 2013. **39**(3): p. 181-5.
151. Taïeb, A. and J.F. Stalder, *Clearing up misunderstandings around core outcomes for atopic dermatitis: reply from the authors.* British Journal of Dermatology, 2015: p. n/a-n/a.
152. Erkut, S., *Developing Multiple Language Versions of Instruments for Intercultural Research.* Child Dev Perspect, 2010. **4**(1): p. 19-24.
153. WHO. *Process of translation and adaptation of instruments*. 30th October 2015]; Available from: http://www.who.int/substance_abuse/research_tools/translation/en/.
154. Wild, D., A. Grove, M. Martin, S. Eremenco, et al., *Principles of Good Practice for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes (PRO) Measures: report of the ISPOR Task Force for Translation and Cultural Adaptation.* Value Health, 2005. **8**(2): p. 94-104.
155. Kirkham, J.J., S. Gorst, D.G. Altman, J. Blazeby, et al., *COS-STAR: a reporting guideline for studies developing core outcome sets (protocol).* Trials, 2015. **16**(1): p. 373.
156. Kirkham, J.J., M. Boers, P. Tugwell, M. Clarke, et al., *Outcome measures in rheumatoid arthritis randomised trials over the last 50 years.* Trials, 2013. **14**(1): p. 324.
157. Barnes, K., M. Clarke, J. Kirkham, and P. Williamson, *Citation analysis: a new approach to assess the uptake of core outcome sets.* Trials 2015. **2015**(16 (Suppl 2)).
158. Bautista-Molano, W., V. Navarro-Compán, R.M. Landewé, M. Boers, et al., *How well are the ASAS/OMERACT Core Outcome Sets for Ankylosing Spondylitis implemented*

in randomized clinical trials? A systematic literature review. Clinical Rheumatology, 2014. **33**(9): p. 1313-1322.
159. *Education section - Studies Within A Trial (SWAT).* J Evid Based Med, 2012. **5**(1): p. 44-5.
160. Brookes, S.T., R.C. Macefield, P.R. Williamson, A.G. McNair, et al., *Three nested RCTs of dual or single stakeholder feedback within Delphi surveys during core outcome and information set development.* Trials 2015. **16(Suppl 2)**.
161. de Wit, M.P., T.K. Kvien, and L. Gossec, *Patient participation as an integral part of patient-reported outcomes development ensures the representation of the patient voice: a case study from the field of rheumatology.* RMD Open, 2015. **1**(1): p. e000129.
162. Snape, D., J. Kirkham, N. Britten, K. Froggatt, et al., *Exploring perceived barriers, drivers, impacts and the need for evaluation of public involvement in health and social care research: a modified Delphi study.* BMJ Open, 2014. **4**(6): p. e004943.
163. Jepson, M., D. Townsend, J. Wade, C. Conefrey, et al., *"It's what we call a randomised control trial" exploring how randomisation is presented by recruiters in RCTs.* Trials, 2015. **16 (Suppl 2)**.
164. Gorst, S.L., E. Gargon, M. Clarke, J.M. Blazeby, et al., *Choosing Important Health Outcomes for Comparative Effectiveness Research: An Updated Review and User Survey.* PLoS ONE, 2016. **11**(1): p. e0146444.
165. NLM. *MeSH Vocabulary Changes* 2015 14/01/2016]; Available from: https://www.nlm.nih.gov/mesh/intro_voc_change.html.
166. Staa, T.P., B. Goldacre, M. Gulliford, J. Cassell, et al., *Pragmatic randomised trials using routine electronic health records: putting them to the test.* BMJ, 2012. **344**: p. e55.
167. Blazeby, J.M., R. Macefield, N.S. Blencowe, M. Jacobs, et al., *Core information set for oesophageal cancer surgery.* Br J Surg, 2015. **102**(8): p. 936-43.
168. Brown, H., A. Hewison, N. Gale, I. Snelling, et al. *Every patient a research patient?* 2015 [cited 2016 10 February]; Available from: https://www.cancerresearchuk.org/sites/default/files/cruk_every_patient_may2015_web.pdf.
169. Davies, S. *We can make research a statement, not a question*. 2015 [cited 2016 10 February]; Available from: http://scienceblog.cancerresearchuk.org/2015/05/19/dame-sally-davies-we-can-make-research-a-statement-not-a-question/.
170. CROWN. *Journals* [cited 2016 10 February]; Available from: http://www.crown-initiative.org/journals/.
171. Dixon, D.O., P. McLaughlin, F.B. Hagemeister, E.J. Freireich, et al., *Reporting outcomes in Hodgkin's disease and lymphoma.* Journal of Clinical Oncology, 1987. **5**(10): p. 1670-2.
172. Glynne-Jones, R., S. Mawdsley, T. Pearce, and M. Buyse, *Alternative clinical end points in rectal cancer--are we getting closer?* Annals of Oncology, 2006. **17**(8): p. 1239-48.
173. Auvinen, A., J.B. Rietbergen, L.J. Denis, F.H. Schroder, et al., *Prospective evaluation plan for randomised trials of prostate cancer screening. The International Prostate Cancer Screening Trial Evaluation Group.* Journal of Medical Screening, 1996. **3**(2): p. 97-104.
174. Denis, L., B.J. Norlén, L. Holmberg, C.B. Begg, et al., *Planning controlled clinical trials.* Urology, 1997. **49**(4 SUPPL.): p. 15-26.
175. Scher, H.I., M. Eisenberger, A.V. D'Amico, S. Halabi, et al., *Eligibility and outcomes reporting guidelines for clinical trials for patients in the state of a rising prostate-*

specific antigen: recommendations from the Prostate-Specific Antigen Working Group.[Erratum appears in J Clin Oncol. 2004 Aug 1;22(15):3205]. Journal of Clinical Oncology, 2004. **22**(3): p. 537-56.
176. Dawson, N.A., *Apples and oranges: building a consensus for standardized eligibility criteria and end points in prostate cancer clinical trials.* Journal of Clinical Oncology, 1998. **16**(10): p. 3398-405.
177. Middleton, R.G., I.M. Thompson, M.S. Austenfeld, W.H. Cooner, et al., *Prostate Cancer Clinical Guidelines Panel Summary report on the management of clinically localized prostate cancer. The American Urological Association.* Journal of Urology, 1995. **154**(6): p. 2144-8.
178. Schellhammer, P., A. Cockett, L. Boccon-Gibod, M. Gospodarowicz, et al., *Assessment of endpoints for clinical trials for localized prostate cancer.* Urology, 1997. **49**(4A Suppl): p. 27-38.
179. Rajkumar, S.V., J.-L. Harousseau, B. Durie, K.C. Anderson, et al., *Consensus recommendations for the uniform reporting of clinical trials: report of the International Myeloma Workshop Consensus Panel 1.* Blood, 2011. **117**(18): p. 4691-5.
180. Party, I.B.M.C.W., *International bone metastases consensus on endpoint measurements for future clinical trials: proceedings of the first survey and meeting (work in progress) International Bone Metastases Consensus Working Party.* Clinical Oncology (Royal College of Radiologists), 2001. **13**(2): p. 82-4.
181. Chow, E., J.S.Y. Wu, P. Hoskin, L.R. Coia, et al., *International consensus on palliative radiotherapy endpoints for future clinical trials in bone metastases.* Radiotherapy & Oncology, 2002. **64**(3): p. 275-80.
182. Partsch, H., N. Stout, I. Forner-Cordero, M. Flour, et al., *Clinical trials needed to evaluate compression therapy in breast cancer related lymphedema (BCRL). Proposals from an expert group.* International Angiology, 2010. **29**(5): p. 442-53.
183. Hesketh, P.J., R.J. Gralla, A. du Bois, and M. Tonato, *Methodology of antiemetic trials: response assessment, evaluation of new agents and definition of chemotherapy emetogenicity.* Supportive Care in Cancer, 1998. **6**(3): p. 221-7.
184. Tonato, M., *Introduction.* Supportive Care in Cancer, 1998. **6**(3): p. 193-193.
185. McVie, J.G. and K.M. de Bruijn, *Methodology of antiemetic trials.* Drugs, 1992. **43 Suppl 3**: p. 1-5.
186. Pallis, A.G., A. Ring, C. Fortpied, B. Penninckx, et al., *EORTC workshop on clinical trial methodology in older individuals with a diagnosis of solid tumors.* Annals of Oncology, 2011. **22**(8): p. 1922-6.
187. Prorok, P.C. and P.M. Marcus, *Cancer screening trials: nuts and bolts.* Seminars in Oncology, 2010. **37**(3): p. 216-23.
188. Punt, C.J.A., M. Buyse, C.-H. Kohne, P. Hohenberger, et al., *Endpoints in adjuvant treatment trials: a systematic review of the literature in colon cancer and proposed definitions for future trials.* Journal of the National Cancer Institute, 2007. **99**(13): p. 998-1003.
189. Wils, J., T. Sahmoud, A. Sobrero, H. Bleiberg, et al., *Evaluation of clinical efficacy of new medical treatments in advanced colorectal cancer. Results of a workshop organized by the EORTC GITCCG. European Organization for Research and Treatment of Cancer. Gastrointestinal Tract Cancer Cooperative Group.* Tumori, 1998. **84**(3): p. 335-47.
190. Llovet, J.M., A.M. Di Bisceglie, J. Bruix, B.S. Kramer, et al., *Design and endpoints of clinical trials in hepatocellular carcinoma.* Journal of the National Cancer Institute, 2008. **100**(10): p. 698-711.

191. Pagliusi, S.R. and M. Teresa Aguado, *Efficacy and other milestones for human papillomavirus vaccine introduction.* Vaccine, 2004. **23**(5): p. 569-78.
192. Lefebvre, J.-L., K.K. Ang, and P. Larynx Preservation Consensus, *Larynx preservation clinical trial design: key issues and recommendations-a consensus panel summary.* International Journal of Radiation Oncology, Biology, Physics, 2009. **73**(5): p. 1293-303.
193. Adelstein, D.J., J.A. Ridge, D.M. Brizel, F.C. Holsinger, et al., *Transoral resection of pharyngeal cancer: summary of a National Cancer Institute Head and Neck Cancer Steering Committee Clinical Trials Planning Meeting, November 6-7, 2011, Arlington, Virginia.* Head & Neck, 2012. **34**(12): p. 1681-703.
194. Gridelli, C., A. Ardizzoni, T. Le Chevalier, C. Manegold, et al., *Treatment of advanced non-small-cell lung cancer patients with ECOG performance status 2: results of an European Experts Panel.* Annals of Oncology, 2004. **15**(3): p. 419-26.
195. Gridelli, C., F. de Marinis, M. Di Maio, A. Ardizzoni, et al., *Maintenance treatment of advanced non-small-cell lung cancer: Results of an International Expert Panel Meeting of the Italian Association of Thoracic Oncology.* Lung Cancer, 2012. **76**(3): p. 269-279.
196. Cheson, B.D., S.J. Horning, B. Coiffier, M.A. Shipp, et al., *Report of an international workshop to standardize response criteria for non-Hodgkin's lymphomas. NCI Sponsored International Working Group.* Journal of Clinical Oncology, 1999. **17**(4): p. 1244.
197. Cheson, B.D., B. Pfistner, M.E. Juweid, R.D. Gascoyne, et al., *Revised response criteria for malignant lymphoma.* Journal of Clinical Oncology, 2007. **25**(5): p. 579-86.
198. Anderson, K.C., R.A. Kyle, S.V. Rajkumar, A.K. Stewart, et al., *Clinically relevant end points and new drug approvals for myeloma.* Leukemia, 2008. **22**(2): p. 231-9.
199. Cheson, B.D., J.M. Bennett, K.J. Kopecky, T. Buchner, et al., *Revised recommendations of the International Working Group for Diagnosis, Standardization of Response Criteria, Treatment Outcomes, and Reporting Standards for Therapeutic Trials in Acute Myeloid Leukemia.[Erratum appears in J Clin Oncol. 2004 Feb 1;22(3):576 Note: LoCocco, Francesco [corrected to Lo-Coco, Francesco]].* Journal of Clinical Oncology, 2003. **21**(24): p. 4642-9.
200. Kulke, M.H., L.L. Siu, J.E. Tepper, G. Fisher, et al., *Future directions in the treatment of neuroendocrine tumors: consensus report of the National Cancer Institute Neuroendocrine Tumor clinical trials planning meeting.* Journal of Clinical Oncology, 2011. **29**(7): p. 934-43.
201. Bellm, L.A., G. Cunningham, L. Durnell, J. Eilers, et al., *Defining clinically meaningful outcomes in the evaluation of new treatments for oral mucositis: oral mucositis patient provider advisory board.* Cancer Investigation, 2002. **20**(5-6): p. 793-800.
202. du Bois, A., M. Quinn, T. Thigpen, J. Vermorken, et al., *2004 consensus statements on the management of ovarian cancer: final document of the 3rd International Gynecologic Cancer Intergroup Ovarian Cancer Consensus Conference (GCIG OCCC 2004).* Annals of Oncology, 2005. **16 Suppl 8**: p. viii7-viii12.
203. Stuart, G.C.E., H. Kitchener, M. Bacon, A. duBois, et al., *2010 Gynecologic Cancer InterGroup (GCIG) consensus statement on clinical trials in ovarian cancer: report from the Fourth Ovarian Cancer Consensus Conference.* International Journal of Gynecological Cancer, 2011. **21**(4): p. 750-5.
204. Thigpen, T., A. duBois, J. McAlpine, P. DiSaia, et al., *First-line therapy in ovarian cancer trials.* International Journal of Gynecological Cancer, 2011. **21**(4): p. 756-62.

205. Comenzo, R.L., D. Reece, G. Palladini, D. Seldin, et al., *Consensus guidelines for the conduct and reporting of clinical trials in systemic light-chain amyloidosis.* Leukemia, 2012. **26**(11): p. 2317-25.
206. Dorman, S., C. Jolley, A. Abernethy, D. Currow, et al., *Researching breathlessness in palliative care: consensus statement of the National Cancer Research Institute Palliative Care Breathlessness Subgroup.* Palliative Medicine, 2009. **23**(3): p. 213-27.
207. Renal Disease Subcommittee of the American College of Rheumatology Ad Hoc Committee on Systemic Lupus Erythematosus Response, C., *The American College of Rheumatology response criteria for proliferative and membranous renal disease in systemic lupus erythematosus clinical trials.* Arthritis & Rheumatism, 2006. **54**(2): p. 421-32.
208. Bertsias, G.K., J.P.A. Ioannidis, J. Boletis, S. Bombardieri, et al., *EULAR points to consider for conducting clinical trials in systemic lupus erythematosus: literature based evidence for the selection of endpoints.* Annals of the Rheumatic Diseases, 2009. **68**(4): p. 477-83.
209. Gordon, C., G. Bertsias, J.P.A. Ioannidis, J. Boletis, et al., *EULAR points to consider for conducting clinical trials in systemic lupus erythematosus.* Annals of the Rheumatic Diseases, 2009. **68**(4): p. 470-6.
210. Smolen, J.S., V. Strand, M. Cardiel, S. Edworthy, et al., *Randomized clinical trials and longitudinal observational studies in systemic lupus erythematosus: consensus on a preliminary core set of outcome domains.* Journal of Rheumatology, 1999. **26**(2): p. 504-7.
211. Ruperto, N., A. Ravelli, K.J. Murray, D.J. Lovell, et al., *Preliminary core sets of measures for disease activity and damage assessment in juvenile systemic lupus erythematosus and juvenile dermatomyositis.* Rheumatology, 2003. **42**(12): p. 1452-9.
212. White, B., E.A. Bauer, L.A. Goldsmith, M.C. Hochberg, et al., *Guidelines for clinical trials in systemic sclerosis (scleroderma). I. Disease-modifying interventions. The American College of Rheumatology Committee on Design and Outcomes in Clinical Trials in Systemic Sclerosis.* Arthritis & Rheumatism, 1995. **38**(3): p. 351-60.
213. Khanna, D., D.J. Lovell, E. Giannini, P.J. Clements, et al., *Development of a provisional core set of response measures for clinical trials of systemic sclerosis.* Annals of the Rheumatic Diseases, 2008. **67**(5): p. 703-9.
214. Clements, P.J., Y. Allanore, D. Khanna, M. Singh, et al., *Arthritis in Systemic Sclerosis: Systematic Review of the Literature and Suggestions for the Performance of Future Clinical Trials in Systemic Sclerosis Arthritis.* Seminars in Arthritis and Rheumatism, 2012. **41**(6): p. 801-814.
215. Khanna, D., K.K. Brown, P.J. Clements, R. Elashoff, et al., *Systemic sclerosis-associated interstitial lung disease-proposed recommendations for future randomized clinical trials.* Clinical & Experimental Rheumatology, 2010. **28**(2 Suppl 58): p. S55-62.
216. Merkel, P.A., K. Herlyn, A.D. Mahr, T. Neogi, et al., *Progress towards a core set of outcome measures in small-vessel vasculitis. Report from OMERACT 9.* Journal of Rheumatology, 2009. **36**(10): p. 2362-8.
217. Merkel, P.A., S.Z. Aydin, M. Boers, H. Direskeneli, et al., *The OMERACT core set of outcome measures for use in clinical trials of ANCA-associated vasculitis.* Journal of Rheumatology, 2011. **38**(7): p. 1480-6.
218. Hellmich, B., O. Flossmann, W.L. Gross, P. Bacon, et al., *EULAR recommendations for conducting clinical studies and/or clinical trials in systemic vasculitis: focus on anti-neutrophil cytoplasm antibody-associated vasculitis.* Annals of the Rheumatic Diseases, 2007. **66**(5): p. 605-17.

219. Felson, D.T., *Choosing a core set of disease activity measures for rheumatoid arthritis clinical trials.* Journal of Rheumatology, 1993. **20**(3): p. 531-4.
220. Fried, B.J., M. Boers, and P.R. Baker, *A method for achieving consensus on rheumatoid arthritis outcome measures: the OMERACT conference process.* Journal of Rheumatology, 1993. **20**(3): p. 548-51.
221. Tugwell, P. and M. Boers, *Developing consensus on preliminary core efficacy endpoints for rheumatoid arthritis clinical trials. OMERACT Committee.* Journal of Rheumatology, 1993. **20**(3): p. 555-6.
222. Sanderson, T., M. Morris, M. Calnan, P. Richards, et al., *Patient perspective of measuring treatment efficacy: the rheumatoid arthritis patient priorities for pharmacologic interventions outcomes.* Arthritis care & research, 2010. **62**(5): p. 647-56.
223. Sanderson, T., M. Morris, M. Calnan, P. Richards, et al., *What outcomes from pharmacologic treatments are important to people with rheumatoid arthritis? Creating the basis of a patient core set.* Arthritis care & research, 2010. **62**(5): p. 640-6.
224. Bombardier, C., P. Tugwell, A. Sinclair, C. Dok, et al., *Preference for endpoint measures in clinical trials: results of structured workshops.* Journal of Rheumatology, 1982. **9**(5): p. 798-801.
225. Scott, D.L., T.D. Spector, T. Pullar, and B. McConkey, *What should we hope to achieve when treating rheumatoid arthritis?* Annals of the Rheumatic Diseases, 1989. **48**(3): p. 256-61.
226. van Riel, P.L., *Provisional guidelines for measuring disease activity in clinical trials on rheumatoid arthritis.* Br J Rheumatol, 1992. **31**(12): p. 793-4.
227. Taylor, W.J., *Preliminary identification of core domains for outcome studies in psoriatic arthritis using Delphi methods.* Annals of the Rheumatic Diseases, 2005. **64 Suppl 2**: p. ii110-2.
228. Gladman, D.D., V. Strand, P.J. Mease, C. Antoni, et al., *OMERACT 7 psoriatic arthritis workshop: synopsis.* Annals of the Rheumatic Diseases, 2005. **64 Suppl 2**: p. ii115-6.
229. Gladman, D.D., *Consensus exercise on domains in psoriatic arthritis.* Annals of the Rheumatic Diseases, 2005. **64 Suppl 2**: p. ii113-4.
230. Gladman, D.D., P.J. Mease, V. Strand, P. Healy, et al., *Consensus on a core set of domains for psoriatic arthritis.* Journal of Rheumatology, 2007. **34**(5): p. 1167-70.
231. Miller, F.W., L.G. Rider, Y.L. Chung, R. Cooper, et al., *Proposed preliminary core set measures for disease outcome assessment in adult and juvenile idiopathic inflammatory myopathies.* Rheumatology, 2001. **40**(11): p. 1262-73.
232. Van Der Heijde, D., N. Bellamy, A. Calin, M. Dougados, et al., *Preliminary core sets for endpoints in ankylosing spondylitis.* Journal of Rheumatology, 1997. **24**(11): p. 2225-2229.
233. Wolfe, F., M. Lassere, D. van der Heijde, G. Stucki, et al., *Preliminary core set of domains and reporting requirements for longitudinal observational studies in rheumatology.* Journal of Rheumatology, 1999. **26**(2): p. 484-9.
234. Mease, P.J., D.J. Clauw, L.M. Arnold, D.L. Goldenberg, et al., *Fibromyalgia syndrome.* Journal of Rheumatology, 2005. **32**(11): p. 2270-7.
235. Mease, P., L.M. Arnold, R. Bennett, A. Boonen, et al., *Fibromyalgia syndrome.* Journal of Rheumatology, 2007. **34**(6): p. 1415-25.
236. Arnold, L.M., L.J. Crofford, P.J. Mease, S.M. Burgess, et al., *Patient perspectives on the impact of fibromyalgia.* Patient Education & Counseling, 2008. **73**(1): p. 114-20.

237. Mease, P.J., L.M. Arnold, L.J. Crofford, D.A. Williams, et al., *Identifying the clinical domains of fibromyalgia: contributions from clinician and patient Delphi exercises.* Arthritis & Rheumatism, 2008. **59**(7): p. 952-60.
238. Carville, S.F., Choy E H, *Systematic review of discriminating power of outcome measures used in clinical trials of fibrymyalgia.* Journal of Rheumatology, 2008. **35**: p. 2094-2105.
239. Choy, E.H., L.M. Arnold, D.J. Clauw, L.J. Crofford, et al., *Content and criterion validity of the preliminary core dataset for clinical trials in fibromyalgia syndrome.* Journal of Rheumatology, 2009. **36**(10): p. 2330-4.
240. Mease, P., L.M. Arnold, E.H. Choy, D.J. Clauw, et al., *Fibromyalgia syndrome module at OMERACT 9: domain construct.* Journal of Rheumatology, 2009. **36**(10): p. 2318-29.
241. Salaffi, F., A. Ciapetti, P. Sarzi Puttini, F. Atzeni, et al., *Preliminary identification of key clinical domains for outcome evaluation in fibromyalgia using the Delphi method: The Italian experience.* Reumatismo, 2012. **64**(1): p. 27-34.
242. Schumacher, H.R., Edwards, N L, Perez-Ruiz, F et al, *Outcome measures for acute and chronic gout.* Journal of Rheumatology, 2005. **32**: p. 2452-5.
243. Schumacher, H.R., W. Taylor, N. Joseph-Ridge, F. Perez-Ruiz, et al., *Outcome evaluations in gout.* Journal of Rheumatology, 2007. **34**(6): p. 1381-5.
244. Taylor, W.J., H.R. Schumacher, Jr., H.S.B. Baraf, P. Chapman, et al., *A modified Delphi exercise to determine the extent of consensus with OMERACT outcome domains for studies of acute and chronic gout.[Erratum appears in Ann Rheum Dis. 2008 Nov;67(11):1652. Note: Mellado, J Vazquez [corrected to Vazquez-Mellado, J]].* Annals of the Rheumatic Diseases, 2008. **67**(6): p. 888-91.
245. Schumacher, H.R., W. Taylor, L. Edwards, R. Grainger, et al., *Outcome domains for studies of acute and chronic gout.* Journal of Rheumatology, 2009. **36**(10): p. 2342-5.
246. Giannini, E.H., N. Ruperto, A. Ravelli, D.J. Lovell, et al., *Preliminary definition of improvement in juvenile arthritis.* Arthritis & Rheumatism, 1997. **40**(7): p. 1202-9.
247. Bellamy, N., J. Kirwan, M. Boers, P. Brooks, et al., *Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III.* Journal of Rheumatology, 1997. **24**(4): p. 799-802.
248. Heiligenhaus, A., I. Foeldvari, C. Edelsten, J.R. Smith, et al., *Proposed outcome measures for prospective clinical trials in juvenile idiopathic arthritis-associated uveitis: a consensus effort from the multinational interdisciplinary working group for uveitis in childhood.* Arthritis care & research, 2012. **64**(9): p. 1365-72.
249. Bowman, S.J., S. Pillemer, R. Jonsson, K. Asmussen, et al., *Revisiting Sjögren's syndrome in the new millennium: Perspectives on assessment and outcome measures. Report of a workshop held on 23 March 2000 at Oxford, UK.* Rheumatology, 2001. **40**(10): p. 1180-1188.
250. Pillemer, S.R., J. Smith, P.C. Fox, and S.J. Bowman, *Outcome measures for Sjogren's syndrome, April 10-11, 2003, Bethesda, Maryland, USA.* Journal of Rheumatology, 2005. **32**(1): p. 143-9.
251. *Guidelines for osteoporosis trials.* Journal of Rheumatology, 1997. **24**(6): p. 1234-6.
252. Sjodahl Hammarlund, C., M.H. Nilsson, and P. Hagell, *Measuring outcomes in Parkinson's disease: a multi-perspective concept mapping study.* Quality of Life Research, 2012. **21**(3): p. 453-63.
253. Vellas, B., S. Andrieu, C. Sampaio, N. Coley, et al., *Endpoints for trials in Alzheimer's disease: a European task force consensus.* Lancet Neurology, 2008. **7**(5): p. 436-50.

254. *World Federation of Neurology Research Group on Neuromuscular Diseases Subcommittee on Motor Neuron Disease. Airlie House guidelines. Therapeutic trials in amyotrophic lateral sclerosis. Airlie House "Therapeutic Trials in ALS" Workshop Contributors.* Journal of the Neurological Sciences, 1995. **129 Suppl**: p. 1-10.
255. Miller, R.G., T.L. Munsat, M. Swash, and B.R. Brooks, *Consensus guidelines for the design and implementation of clinical trials in ALS. World Federation of Neurology committee on Research.* Journal of the Neurological Sciences, 1999. **169**(1-2): p. 2-12.
256. Leigh, P.N., M. Swash, Y. Iwasaki, A. Ludolph, et al., *Amyotrophic lateral sclerosis: a consensus viewpoint on designing and implementing a clinical trial.* Amyotrophic Lateral Sclerosis & Other Motor Neuron Disorders, 2004. **5**(2): p. 84-98.
257. Shankaran, S. and A. Laptook, *Challenge of conducting trials of neuroprotection in the asphyxiated term infant.* Seminars in Perinatology, 2003. **27**(4): p. 320-32.
258. Vargus-Adams, J.N. and L.K. Martin, *Measuring what matters in cerebral palsy: a breadth of important domains and outcome measures.* Archives of Physical Medicine & Rehabilitation, 2009. **90**(12): p. 2089-95.
259. Katona, C., G. Livingston, C. Cooper, D. Ames, et al., *International Psychogeriatric Association consensus statement on defining and measuring treatment benefits in dementia.* International Psychogeriatrics, 2007. **19**(3): p. 345-54.
260. Moniz-Cook, E., M. Vernooij-Dassen, R. Woods, F. Verhey, et al., *A European consensus on outcome measures for psychosocial intervention research in dementia care.* Aging & Mental Health, 2008. **12**(1): p. 14-29.
261. *Considerations on designing clinical trials to evaluate the place of new antiepileptic drugs in the treatment of newly diagnosed and chronic patients with epilepsy.* Epilepsia, 1998. **39**(7): p. 799-803.
262. LaFrance, W.C., Jr., K. Alper, D. Babcock, J.J. Barry, et al., *Nonepileptic seizures treatment workshop summary.* Epilepsy & Behavior, 2006. **8**(3): p. 451-61.
263. Osborne, J.P. and A. Lux, *Towards an international consensus on definitions and standardised outcome measures for therapeutic trials (and epidemiological studies) in West syndrome.* Brain & Development, 2001. **23**(7): p. 677-82.
264. Lux, A.L. and J.P. Osborne, *A proposal for case definitions and outcome measures in studies of infantile spasms and West syndrome: consensus statement of the West Delphi group.* Epilepsia, 2004. **45**(11): p. 1416-28.
265. Mindell, J.A., G. Emslie, J. Blumer, M. Genel, et al., *Pharmacologic management of insomnia in children and adolescents: consensus statement.* Pediatrics, 2006. **117**(6): p. e1223-32.
266. Schumacher, H.C., P.M. Meyers, R.T. Higashida, C.P. Derdeyn, et al., *Reporting standards for angioplasty and stent-assisted angioplasty for intracranial atherosclerosis.* Journal of Neurointerventional Surgery, 2010. **2**(4): p. 324-40.
267. Whitaker, J.N., H.F. McFarland, P. Rudge, and S.C. Reingold, *Outcomes assessment in multiple sclerosis clinical trials: a critical analysis.* Multiple Sclerosis, 1995. **1**(1): p. 37-47.
268. Chitnis, T., S. Tenembaum, B. Banwell, L. Krupp, et al., *Consensus statement: evaluation of new and existing therapeutics for pediatric multiple sclerosis.* Multiple Sclerosis, 2012. **18**(1): p. 116-27.
269. Chitnis, T., M. Tardieu, M.P. Amato, B. Banwell, et al., *International Pediatric MS Study Group Clinical Trials Summit: meeting report.* Neurology, 2013. **80**(12): p. 1161-8.
270. Penzien, D.B., *Guidelines for trials of behavioral treatments for recurrent headache: purpose, process, and product.* Headache, 2005. **45 Suppl 2**: p. S87-9.

271. Penzien, D.B., F. Andrasik, B.M. Freidenberg, T.T. Houle, et al., *Guidelines for trials of behavioral treatments for recurrent headache, first edition: American Headache Society Behavioral Clinical Trials Workgroup.* Headache, 2005. **45 Suppl 2**: p. S110-32.
272. Andrasik, F., G.L. Lipchik, D.C. McCrory, and D.A. Wittrock, *Outcome measurement in behavioral headache research: headache parameters and psychosocial outcomes.* Headache, 2005. **45**(5): p. 429-37.
273. *Guidelines for controlled trials of drugs in migraine. First edition. International Headache Society Committee on Clinical Trials in Migraine.* Cephalalgia, 1991. **11**(1): p. 1-12.
274. Tfelt-Hansen, P., G. Block, C. Dahlof, H.C. Diener, et al., *Guidelines for controlled trials of drugs in migraine: second edition.* Cephalalgia, 2000. **20**(9): p. 765-86.
275. Tfelt-Hansen, P., J. Pascual, N. Ramadan, C. Dahlof, et al., *Guidelines for controlled trials of drugs in migraine: third edition. A guide for investigators.* Cephalalgia, 2012. **32**(1): p. 6-38.
276. Lipton, R.B., G. Micieli, D. Russell, S. Solomon, et al., *Guidelines for controlled trials of drugs in cluster headache.* Cephalalgia, 1995. **15**(6): p. 452-62.
277. Schoenen, J., *Guidelines for trials of drug treatments in tension-type headache. First edition: International Headache Society Committee on Clinical Trials.* Cephalalgia, 1995. **15**(3): p. 165-79.
278. Bendtsen, L., M.E. Bigal, R. Cerbo, H.C. Diener, et al., *Guidelines for controlled trials of drugs in tension-type headache: second edition.* Cephalalgia, 2010. **30**(1): p. 1-16.
279. Hughes, R., Cidp, and M.M.N.t. group, *129th ENMC International Workshop: Clinical Trials for Chronic Inflammatory Demyelinating Polyradiculoneuropathy and Multifocal Motor Neuropathy, 27th October 2004, Schiphol airport, The Netherlands.* Neuromuscular Disorders, 2005. **15**(4): p. 321-5.
280. Merkies, I.S.J. and G. Lauria, *131st ENMC international workshop: selection of outcome measures for peripheral neuropathy clinical trials 10-12 December 2004, Naarden, The Netherlands.* Neuromuscular Disorders, 2006. **16**(2): p. 149-56.
281. Reilly, M.M., P. de Jonghe, and D. Pareyson, *136th ENMC International Workshop: Charcot-Marie-Tooth disease type 1A (CMT1A)8-10 April 2005, Naarden, The Netherlands.* Neuromuscular Disorders, 2006. **16**(6): p. 396-402.
282. Clifton, G.L., R.L. Hayes, H.S. Levin, M.E. Michel, et al., *Outcome measures for clinical trials involving traumatically brain-injured patients: report of a conference.* Neurosurgery, 1992. **31**(5): p. 975-8.
283. Wilde, E.A., G.G. Whiteneck, J. Bogner, T. Bushnik, et al., *Recommendations for the use of common outcome measures in traumatic brain injury research.* Archives of Physical Medicine & Rehabilitation, 2010. **91**(11): p. 1650-1660 e17.
284. Duncan, P.W., H.S. Jorgensen, and D.T. Wade, *Outcome measures in acute stroke trials: a systematic review and some recommendations to improve practice.* Stroke, 2000. **31**(6): p. 1429-38.
285. Schellinger, P.D., P.M.W. Bath, K.R. Lees, N.M. Bornstein, et al., *Assessment of additional endpoints for trials in acute stroke - what, when, where, in who?* International Journal of Stroke, 2012. **7**(3): p. 227-30.
286. Hoeper, M.M., R.J. Oudiz, A. Peacock, V.F. Tapson, et al., *End points and clinical trial designs in pulmonary arterial hypertension: clinical and regulatory perspectives.* Journal of the American College of Cardiology, 2004. **43**(12 Suppl S): p. 48S-55S.
287. Distler, O., F. Behrens, D. Pittrow, D. Huscher, et al., *Defining appropriate outcome measures in pulmonary arterial hypertension related to systemic sclerosis: a Delphi*

consensus study with cluster analysis.[Erratum appears in Arthritis Rheum. 2008 Aug 15;59(8):1202]. Arthritis & Rheumatism, 2008. **59**(6): p. 867-75.
288. Becker, L.B., T.P. Aufderheide, R.G. Geocadin, C.W. Callaway, et al., *Primary outcomes for resuscitation science studies: a consensus statement from the American Heart Association.* Circulation, 2011. **124**(19): p. 2158-77.
289. Chiam, P.T.L. and C.E. Ruiz, *Percutaneous transcatheter aortic valve implantation: assessing results, judging outcomes, and planning trials: the interventionalist perspective.* Jacc: Cardiovascular Interventions, 2008. **1**(4): p. 341-50.
290. Leon, M.B., N. Piazza, E. Nikolsky, E.H. Blackstone, et al., *Standardized endpoint definitions for transcatheter aortic valve implantation clinical trials: a consensus report from the Valve Academic Research Consortium.* European Heart Journal, 2011. **32**(2): p. 205-17.
291. Kappetein, A.P., S.J. Head, P. Genereux, N. Piazza, et al., *Updated standardized endpoint definitions for transcatheter aortic valve implantation: the Valve Academic Research Consortium-2 consensus document (VARC-2).* European Journal of Cardio-Thoracic Surgery, 2012. **42**(5): p. S45-60.
292. Cutlip, D.E., S. Windecker, R. Mehran, A. Boam, et al., *Clinical end points in coronary stent trials: a case for standardized definitions.* Circulation, 2007. **115**(17): p. 2344-51.
293. Simons, M., R.O. Bonow, N.A. Chronos, D.J. Cohen, et al., *Clinical trials in coronary angiogenesis: issues, problems, consensus: An expert panel summary.* Circulation, 2000. **102**(11): p. E73-86.
294. Conte, M.S., P.J. Geraghty, A.W. Bradbury, N.D. Hevelone, et al., *Suggested objective performance goals and clinical trial design for evaluating catheter-based treatment of critical limb ischemia.* Journal of Vascular Surgery, 2009. **50**(6): p. 1462-73.e1-3.
295. Timaran, C.H., J.F. McKinsey, P.A. Schneider, and F. Littooy, *Reporting standards for carotid interventions from the Society for Vascular Surgery.* Journal of Vascular Surgery, 2011. **53**(6): p. 1679-95.
296. Stout, N., H. Partsch, G. Szolnoky, I. Forner-Cordero, et al., *Chronic edema of the lower extremities: international consensus recommendations for compression therapy clinical research trials.* International Angiology, 2012. **31**(4): p. 316-29.
297. Nedeltchev, K., P.M. Pattynama, G. Biaminoo, N. Diehm, et al., *Standardized definitions and clinical endpoints in carotid artery and supra-aortic trunk revascularization trials.* Catheterization & Cardiovascular Interventions, 2010. **76**(3): p. 333-44.
298. Anderson, H.V., W.S. Weintraub, M.J. Radford, M.S. Kremers, et al., *Standardized cardiovascular data for clinical research, registries, and patient care: A report from the data standards workgroup of the national cardiovascular research infrastructure project.* Journal of the American College of Cardiology, 2013. **61**(18): p. 1835-1846.
299. Hausenloy, D.J., H. Erik Botker, G. Condorelli, P. Ferdinandy, et al., *Translating cardioprotection for patient benefit: Position paper from the Working Group of Cellular Biology of the Heart of the European Society of Cardiology.* Cardiovascular Research, 2013. **98**(1): p. 7-27.
300. Labs, K.H., J.A. Dormandy, K.A. Jaeger, C. Stuerzebecher, et al., *Trans-atlantic conference on clinical trial guidelines in PAOD (Peripheral arterial occlusive disease) clinical trial methodology.* European Journal of Vascular & Endovascular Surgery, 1999. **18**(3): p. 253-65.
301. Mitchell, L.G., N.A. Goldenberg, C. Male, G. Kenet, et al., *Definition of clinical efficacy and safety outcomes for clinical trials in deep venous thrombosis and*

pulmonary embolism in children. Journal of Thrombosis & Haemostasis, 2011. **9**(9): p. 1856-8.
302. Steg, P.G., K. Huber, F. Andreotti, H. Arnesen, et al., *Bleeding in acute coronary syndromes and percutaneous coronary interventions: position paper by the Working Group on Thrombosis of the European Society of Cardiology.* European Heart Journal, 2011. **32**(15): p. 1854-64.
303. O'Connell, J.B., P.M. McCarthy, G. Sopko, G.S. Filippatos, et al., *Mechanical circulatory support devices for acute heart failure syndromes: considerations for clinical trial design.* Heart Failure Reviews, 2009. **14**(2): p. 101-12.
304. Higashida, R.T., A.J. Furlan, H. Roberts, T. Tomsick, et al., *Trial design and reporting standards for intra-arterial cerebral thrombolysis for acute ischemic stroke.[Erratum appears in Stroke. 2003 Nov;34(11):2774].* Stroke, 2003. **34**(8): p. e109-37.
305. Calkins, H., K.H. Kuck, R. Cappato, J. Brugada, et al., *2012 HRS/EHRA/ECAS expert consensus statement on catheter and surgical ablation of atrial fibrillation: recommendations for patient selection, procedural techniques, patient management and follow-up, definitions, endpoints, and research trial design.* Journal of Interventional Cardiac Electrophysiology, 2012. **33**(2): p. 171-257.
306. Kirchhof, P., A. Auricchio, J. Bax, H. Crijns, et al., *Outcome parameters for trials in atrial fibrillation: Executive summary - Recommendations from a consensus conference organized by the German Atrial Fibrillation Competence NETwork (AFNET) and the European Heart Rhythm Association (EHRA).* European Heart Journal, 2007. **28**(22): p. 2803-2817.
307. Buser, D.A. and M. Tonetti, *Clinical trials on implants in regenerated bone.* Annals of Periodontology, 1997. **2**(1): p. 329-42.
308. Page, R.C. and T.A. DeRouen, *Design issues specific to studies of periodontitis.* Journal of Periodontal Research, 1992. **27**(4 Pt 2): p. 395-404; discussion 412-6.
309. Imrey, P.B., N.W. Chilton, B.L. Pihlstrom, H.M. Proskin, et al., *Proposed guidelines for American Dental Association acceptance of products for professional, non-surgical treatment of adult periodontitis. Task Force on Design and Analysis in Dental and Oral Research.* Journal of Periodontal Research, 1994. **29**(5): p. 348-60.
310. Lightfoot, W.S., A. Hefti, and A. Mariotti, *Using a Delphi panel to survey criteria for successful periodontal therapy in anterior teeth.* Journal of Periodontology, 2005. **76**(9): p. 1508-12.
311. Lightfoot, W.S., A. Hefti, and A. Mariotti, *Using a Delphi panel to survey criteria for successful periodontal therapy in posterior teeth.* Journal of Periodontology, 2005. **76**(9): p. 1502-7.
312. Weber, H.P., J.P. Fiorellini, and M.C. Jeffcoat, *Clinical trials on placement of implants in existing bone.* Annals of Periodontology, 1997. **2**(1): p. 315-28.
313. Cochran, D.L., *Report of the American Academy of Periodontology's Workshop on the Design and Conduct of Clinical Trials for Endosseous Dental Implants.* Journal of Periodontology, 1998. **69**(2): p. 280-4.
314. Tonetti, M. and R. Palmer, *Clinical research in implant dentistry: Study design, reporting and outcome measurements: Consensus report of Working Group 2 of the VIII European Workshop on Periodontology.* Journal of Clinical Periodontology, 2012. **39**(SUPPL.12): p. 73-80.
315. Smaïl-Faugeron, V., H. Fron Chabouis, P. Durieux, J.P. Attal, et al., *Development of a Core Set of Outcomes for Randomized Controlled Trials with Multiple Outcomes - Example of Pulp Treatments of Primary Teeth for Extensive Decay in Children.* PLoS ONE, 2013. **8**(1).

316. Chilton, N.W. and J.L. Fleiss, *Design and analysis of plaque and gingivitis clinical trials.* J Clin Periodontol, 1986. **13**(5): p. 400-10.
317. *Guidelines for acceptance of chemotherapeutic products for the control of supragingival dental plaque and gingivitis. Council on Dental Therapeutics.* J Am Dent Assoc, 1986. **112**(4): p. 529-32.
318. Pitts, N.B. and J.W. Stamm, *International Consensus Workshop on Caries Clinical Trials (ICW-CCT)--final consensus statements: agreeing where the evidence leads.* J Dent Res, 2004. **83 Spec No C**: p. C125-8.
319. Marshall, J.C., J.-L. Vincent, G. Guyatt, D.C. Angus, et al., *Outcome measures for clinical research in sepsis: a report of the 2nd Cambridge Colloquium of the International Sepsis Forum.* Critical Care Medicine, 2005. **33**(8): p. 1708-16.
320. Goldstein, B., B. Giroir, A. Randolph, and S. International Consensus Conference on Pediatric, *International pediatric sepsis consensus conference: definitions for sepsis and organ dysfunction in pediatrics.* Pediatric Critical Care Medicine, 2005. **6**(1): p. 2-8.
321. Wood, M.J., H. Balfour, K. Beutner, J. Bruxelle, et al., *How should zoster trials be conducted?* Journal of Antimicrobial Chemotherapy, 1995. **36**(6): p. 1089-1101.
322. Barlow, G.D., D.L. Lamping, P.G. Davey, and D. Nathwani, *Evaluation of outcomes in community-acquired pneumonia: a guide for patients, physicians, and policy-makers.* The Lancet Infectious Diseases, 2003. **3**(8): p. 476-88.
323. Spellberg, B., G.H. Talbot, E.P. Brass, J.S. Bradley, et al., *Position paper: recommended design features of future clinical trials of antibacterial agents for community-acquired pneumonia.* Clinical Infectious Diseases, 2008. **47 Suppl 3**: p. S249-65.
324. Powers, J.H., *Recommendations for improving the design, conduct, and analysis of clinical trials in hospital-acquired pneumonia and ventilator-associated pneumonia.* Clinical Infectious Diseases, 2010. **51 Suppl 1**: p. S18-28.
325. Spellberg, B., G. Talbot, A. Infectious Diseases Society of, P. American College of Chest, et al., *Recommended design features of future clinical trials of antibacterial agents for hospital-acquired bacterial pneumonia and ventilator-associated bacterial pneumonia.* Clinical Infectious Diseases, 2010. **51 Suppl 1**: p. S150-70.
326. McCracken, G.H., M.A. Sande, A. Lentnek, R.J. Whitley, et al., *Evaluation of new anti-infective drugs for the treatment of acute bacterial meningitis. Infectious Diseases Society of America and the Food and Drug Administration.* Clinical Infectious Diseases, 1992. **15 Suppl 1**: p. S182-8.
327. Alioum, A., F. Dabis, L. Dequae-Merchadou, G. Haverkamp, et al., *Estimating the efficacy of interventions to prevent mother-to-child transmission of HIV in breast-feeding population: Development of a consensus methodology.* Statistics in Medicine, 2001. **20**(23): p. 3539-3556.
328. Kirkby, R., C. Calabrese, L. Kaltman, J. Monnier, et al., *Methodological considerations for future controlled influenza treatment trials in complementary and alternative medicine.* Journal of Alternative & Complementary Medicine, 2010. **16**(3): p. 275-83.
329. Nystrom, P.O., R. Bax, E.P. Dellinger, L. Dominioni, et al., *Proposed definitions for diagnosis, severity scoring, stratification, and outcome for trials on intraabdominal infection. Joint Working Party of SIS North America and Europe.* World Journal of Surgery, 1990. **14**(2): p. 148-58.
330. Cross, H., *A Delphi consensus on criteria for contraindications, assessment indicators and expected outcomes related to tibialis posterior transfer surgery.* International Journal of Leprosy & Other Mycobacterial Diseases, 2005. **73**(1): p. 13-21.

331. Moorthy, V., Z. Reed, P.G. Smith, and W.H.O.S.G.o.M.o.M.V. Efficacy, *Measurement of malaria vaccine efficacy in phase III trials: report of a WHO consultation.* Vaccine, 2007. **25**(28): p. 5115-23.
332. Moorthy, V.S., Z. Reed, P.G. Smith, and W.H.O.M.V.A. Committee, *MALVAC 2008: Measures of efficacy of malaria vaccines in phase 2b and phase 3 trials--scientific, regulatory and public health perspectives.* Vaccine, 2009. **27**(5): p. 624-8.
333. Steeves, J.D., D. Lammertse, A. Curt, J.W. Fawcett, et al., *Guidelines for the conduct of clinical trials for spinal cord injury (SCI) as developed by the ICCP panel: clinical trial outcome measures.* Spinal Cord, 2007. **45**(3): p. 206-21.
334. Bombardier, C., *Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations.* Spine, 2000. **25**(24): p. 3100-3.
335. Deyo, R.A., M. Battie, A.J. Beurskens, C. Bombardier, et al., *Outcome measures for low back pain research. A proposal for standardized use.* Spine, 1998. **23**(18): p. 2003-13.
336. Devogelaer, J.P., R.L. Dreiser, E. Abadie, B. Avouac, et al., *Guidelines for clinical studies assessing the efficacy of drugs for the management of acute low back pain.* Clinical & Experimental Rheumatology, 2003. **21**(6): p. 691-4.
337. Goldhahn, J., W.H. Scheele, B.H. Mitlak, E. Abadie, et al., *Clinical evaluation of medicinal products for acceleration of fracture healing in patients with osteoporosis.* Bone, 2008. **43**(2): p. 343-7.
338. Lynch, A.D., D.S. Logerstedt, H. Grindem, I. Eitzen, et al., *Consensus criteria for defining 'successful outcome' after ACL injury and reconstruction: A Delaware-Oslo ACL cohort investigation.* British Journal of Sports Medicine, 2013.
339. Falder, S., A. Browne, D. Edgar, E. Staples, et al., *Core outcomes for adult burn survivors: a clinical overview.* Burns, 2009. **35**(5): p. 618-41.
340. Reneman, M.F., T.T. Beemster, M.J.A. Edelaar, J.M. van Velzen, et al., *Towards an ICF- and IMMPACT-Based Pain Vocational Rehabilitation Core Set in the Netherlands.* Journal of Occupational Rehabilitation, 2013: p. 1-9.
341. Lamb, S.E., E.C. Jorstad-Stein, K. Hauer, C. Becker, et al., *Development of a common outcome data set for fall injury prevention trials: the Prevention of Falls Network Europe consensus.* Journal of the American Geriatrics Society, 2005. **53**(9): p. 1618-22.
342. Cameron, I.D., S. Robinovitch, S. Birge, P. Kannus, et al., *Hip protectors: recommendations for conducting clinical trials--an international consensus statement (part II).* Osteoporosis International, 2010. **21**(1): p. 1-10.
343. Smith, M.A., S.R. Leeder, B. Jalaludin, and W.T. Smith, *The asthma health outcome indicators study.* Australian & New Zealand Journal of Public Health, 1996. **20**(1): p. 69-75.
344. Fuhlbrigge, A., D. Peden, A.J. Apter, H.A. Boushey, et al., *Asthma outcomes: exacerbations.* Journal of Allergy & Clinical Immunology, 2012. **129**(3 Suppl): p. S34-48.
345. Akinbami, L.J., S.D. Sullivan, J.D. Campbell, R.W. Grundmeier, et al., *Asthma outcomes: healthcare utilization and costs.* Journal of Allergy & Clinical Immunology, 2012. **129**(3 Suppl): p. S49-64.
346. Szefler, S.J., S. Wenzel, R. Brown, S.C. Erzurum, et al., *Asthma outcomes: biomarkers.* Journal of Allergy & Clinical Immunology, 2012. **129**(3 Suppl): p. S9-23.
347. Tepper, R.S., R.S. Wise, R. Covar, C.G. Irvin, et al., *Asthma outcomes: pulmonary physiology.* Journal of Allergy & Clinical Immunology, 2012. **129**(3 Suppl): p. S65-87.
348. Cloutier, M.M., M. Schatz, M. Castro, N. Clark, et al., *Asthma outcomes: Composite scores of asthma control.* Journal of Allergy and Clinical Immunology, 2012. **129**(3, Supplement): p. S24-S33.

349. Krishnan, J.A., R.F. Lemanske Jr, G.J. Canino, K.S. Elward, et al., *Asthma outcomes: Symptoms.* Journal of Allergy and Clinical Immunology, 2012. **129**(3, Supplement): p. S124-S135.
350. Wilson, S.R., C.S. Rand, M.D. Cabana, M.B. Foggs, et al., *Asthma outcomes: Quality of life.* Journal of Allergy and Clinical Immunology, 2012. **129**(3, Supplement): p. S88-S123.
351. Keim, S.M., D.W. Spaite, R.F. Maio, H.G. Garrison, et al., *Risk adjustment and outcome measures for out-of-hospital respiratory distress.* Academic Emergency Medicine, 2004. **11**(10): p. 1074-81.
352. Canonica, G.W., C.E. Baena-Cagnani, J. Bousquet, P.J. Bousquet, et al., *Recommendations for standardization of clinical trials with Allergen Specific Immunotherapy for respiratory allergy. A statement of a World Allergy Organization (WAO) taskforce.* Allergy, 2007. **62**(3): p. 317-24.
353. Cazzola, M., W. MacNee, F.J. Martinez, K.F. Rabe, et al., *Outcomes for COPD pharmacological trials: from lung function to biomarkers.* European Respiratory Journal, 2008. **31**(2): p. 416-69.
354. *Recommendations for guidelines on clinical trials of mucoactive drugs in chronic bronchitis and chronic obstructive pulmonary disease. Task Group on Mucoactive Drugs.* Chest, 1994. **106**(5): p. 1532-7.
355. Dent, J., P.J. Kahrilas, N. Vakil, S.V. Van Zanten, et al., *Clinical trial design in adult reflux disease: a methodological workshop.* Alimentary Pharmacology & Therapeutics, 2008. **28**(1): p. 107-26.
356. Wirth, S., D. Kelly, E. Sokal, P. Socha, et al., *Guidance for clinical trials for children and adolescents with chronic hepatitis C.* Journal of Pediatric Gastroenterology & Nutrition, 2011. **52**(2): p. 233-7.
357. Bajaj, J.S., J. Cordoba, K.D. Mullen, P. Amodio, et al., *Review article: the design of clinical trials in hepatic encephalopathy--an International Society for Hepatic Encephalopathy and Nitrogen Metabolism (ISHEN) consensus statement.* Alimentary Pharmacology & Therapeutics, 2011. **33**(7): p. 739-47.
358. Fekety, R., H.L. DuPont, M. Cooperstock, M.L. Corrado, et al., *Evaluation of new anti-infective drugs for the treatment of antibiotic-associated colitis. Infectious Diseases Society of America and the Food and Drug Administration.* Clinical Infectious Diseases, 1992. **15 Suppl 1**: p. S263-7.
359. Griffiths, A.M., A.R. Otley, J. Hyams, A.R. Quiros, et al., *A review of activity indices and end points for clinical trials in children with Crohn's disease.* Inflammatory Bowel Diseases, 2005. **11**(2): p. 185-96.
360. Laine, L., B. Spiegel, A. Rostom, P. Moayyedi, et al., *Methodology for randomized trials of patients with nonvariceal upper gastrointestinal bleeding: recommendations from an international consensus conference.* American Journal of Gastroenterology, 2010. **105**(3): p. 540-50.
361. Pimentel, M., N.J. Talley, E.M. Quigley, A. Hani, et al., *Report from the multinational irritable bowel syndrome initiative 2012.* Gastroenterology, 2013. **144**(7): p. e1-5.
362. Sanyal, A.J., E.M. Brunt, D.E. Kleiner, K.V. Kowdley, et al., *Endpoints and clinical trial design for nonalcoholic steatohepatitis.* Hepatology, 2011. **54**(1): p. 344-53.
363. Rahn, D.D., H. Abed, V.W. Sung, K.A. Matteson, et al., *Systematic review highlights difficulty interpreting diverse clinical outcomes in abnormal uterine bleeding trials.* Journal of Clinical Epidemiology, 2011. **64**(3): p. 293-300.
364. Meuleman, C., C. Tomassetti, and T.M. D'Hooghe, *Clinical outcome after laparoscopic radical excision of endometriosis and laparoscopic segmental bowel resection.* Current Opinion in Obstetrics & Gynecology, 2012. **24**(4): p. 245-52.

365. Vincent, K., S. Kennedy, and P. Stratton, *Pain scoring in endometriosis: entry criteria and outcome measures for clinical trials. Report from the Art and Science of Endometriosis meeting.* Fertility & Sterility, 2010. **93**(1): p. 62-7.
366. Broder, M.S., W.J. Landow, S.C. Goodwin, R.H. Brook, et al., *An agenda for research into uterine artery embolization: results of an expert panel conference.* Journal of Vascular & Interventional Radiology, 2000. **11**(4): p. 509-15.
367. Basson, R., J. Berman, A. Burnett, L. Derogatis, et al., *Report of the international consensus development conference on female sexual dysfunction: definitions and classifications.* Journal of Urology, 2000. **163**(3): p. 888-93.
368. Clayton, A.H., L. Dennerstein, W.A. Fisher, S.A. Kingsberg, et al., *Standards for clinical trials in sexual dysfunction in women: research designs and outcomes assessment.* Journal of Sexual Medicine, 2010. **7**(1 Pt 2): p. 541-60.
369. Walker, M., T. Toneatto, M.N. Potenza, N. Petry, et al., *A framework for reporting outcomes in problem gambling treatment research: the Banff, Alberta Consensus.* Addiction, 2006. **101**(4): p. 504-11.
370. Del Boca, F.K. and J. Darkes, *Enhancing the validity and utility of randomized clinical trials in addictions treatment research: II. Participant samples and assessment.* Addiction, 2007. **102**(8): p. 1194-203.
371. Donovan, D.M., G.E. Bigelow, G.S. Brigham, K.M. Carroll, et al., *Primary outcome indices in illicit drug dependence treatment research: systematic approach to selection and measurement of drug use end-points in clinical trials.* Addiction, 2012. **107**(4): p. 694-708.
372. Vocci, F. and H.d. Wit, *Consensus statement on evaluation of outcome of pharmacotherapy for substance abuse/ dependence: report from a NIDA/CPDD meeting.* National Institute on Drug Abuse Medications Development Division, 1999.
373. Levine, L.A. and J.M. Greenfield, *Establishing a standardized evaluation of the man with Peyronie's disease.* International Journal of Impotence Research, 2003. **15 Suppl 5**: p. S103-12.
374. Djurhuus, J.C., J.P. Norgaard, and K. Hjalmas, *What is an acceptable treatment outcome?* Scandinavian Journal of Urology & Nephrology Supplementum, 1997. **183**: p. 75-7.
375. Toozs-Hobson, P., R. Freeman, M. Barber, C. Maher, et al., *An International Urogynecological Association (IUGA)/International Continence Society (ICS) joint report on the terminology for reporting outcomes of surgical procedures for pelvic organ prolapse.* International Urogynecology Journal, 2012. **23**(5): p. 527-35.
376. Porst, H., Y. Vardi, E. Akkus, A. Melman, et al., *Standards for clinical trials in male sexual dysfunctions.* Journal of Sexual Medicine, 2010. **7**(1 Pt 2): p. 414-44.
377. Pavletic, S.Z., P. Martin, S.J. Lee, S. Mitchell, et al., *Measuring therapeutic response in chronic graft-versus-host disease: National Institutes of Health Consensus Development Project on Criteria for Clinical Trials in Chronic Graft-versus-Host Disease: IV. Response Criteria Working Group report.* Biology of Blood & Marrow Transplantation, 2006. **12**(3): p. 252-66.
378. Lassila, R., C. Rothschild, P. De Moerloose, M. Richards, et al., *Recommendations for postmarketing surveillance studies in haemophilia and other bleeding disorders.* Haemophilia, 2005. **11**(4): p. 353-9.
379. Rodeghiero, F., R. Stasi, T. Gernsheimer, M. Michel, et al., *Standardization of terminology, definitions and outcome criteria in immune thrombocytopenic purpura of adults and children: report from an international working group.* Blood, 2009. **113**(11): p. 2386-93.

380. Turk, D.C., R.H. Dworkin, D. Revicki, G. Harding, et al., *Identifying important outcome domains for chronic pain clinical trials: an IMMPACT survey of people with pain.* Pain, 2008. **137**(2): p. 276-85.
381. Apfel, C., L. Eberhart, P. Kranke, and D. Ruesch, *Recommendations for randomized controlled trials to prevent or treat postoperative nausea and vomiting.* Anaesth Intensivmed Notfallmed Schmerzther 2002. **43**: p. 69-74.
382. Anderson, J.W., F.X. Pi-Sunyer, E. Danforth, C.A. Dujovne, et al., *Clinical trial design for obesity agents: a workshop report.* Obesity Research, 1998. **6**(4): p. 311-5.
383. Douglas, R.S., A. Tsirbas, M. Gordon, D. Lee, et al., *Development of criteria for evaluating clinical response in thyroid eye disease using a modified Delphi technique.* Archives of Ophthalmology, 2009. **127**(9): p. 1155-60.
384. Carlson, G.A., P.S. Jensen, R.L. Findling, R.E. Meyer, et al., *Methodological issues and controversies in clinical trials with child and adolescent patients with bipolar disorder: report of a consensus conference.* Journal of Child & Adolescent Psychopharmacology, 2003. **13**(1): p. 13-27.
385. Rush, A.J., H.C. Kraemer, H.A. Sackeim, M. Fava, et al., *Report by the ACNP Task Force on response and remission in major depressive disorder.* Neuropsychopharmacology, 2006. **31**(9): p. 1841-53.
386. Fitzpatrick, R., J. Chambers, T. Burns, H. Doll, et al., *A systematic review of outcome measures used in forensic mental health research with consensus panel opinion.* Health Technology Assessment (Winchester, England), 2010. **14**(18): p. 1-94.
387. Finer, N.N., R. Higgins, J. Kattwinkel, and R.J. Martin, *Summary proceedings from the apnea-of-prematurity group.* Pediatrics, 2006. **117**(3 Pt 2): p. S47-51.
388. Short, B.L., K. Van Meurs, J.R. Evans, and G. Cardiology, *Summary proceedings from the cardiology group on cardiovascular instability in preterm infants.* Pediatrics, 2006. **117**(3 Pt 2): p. S34-9.
389. Clancy, R.R., *Summary proceedings from the neurology group on neonatal seizures.* Pediatrics, 2006. **117**(3 pt. 2): p. S23-S27.
390. Gonzalez, U., M. Whitton, V. Eleftheriadou, M. Pinart, et al., *Guidelines for designing and reporting clinical trials in vitiligo.* Archives of Dermatology, 2011. **147**(12): p. 1428-36.
391. Eleftheriadou, V., K.S. Thomas, M.E. Whitton, J.M. Batchelor, et al., *Which outcomes should we measure in vitiligo? Results of a systematic review and a survey among patients and clinicians on outcomes in vitiligo trials.* British Journal of Dermatology, 2012. **167**(4): p. 804-14.
392. Schmitt, J., S. Langan, H.C. Williams, and N. European Dermato-Epidemiology, *What are the best outcome measurements for atopic eczema? A systematic review.* Journal of Allergy & Clinical Immunology, 2007. **120**(6): p. 1389-98.
393. Schmitt, J., H. Williams, and H.D. Group, *Harmonising Outcome Measures for Eczema (HOME). Report from the First International Consensus Meeting (HOME 1), 24 July 2010, Munich, Germany.* British Journal of Dermatology, 2010. **163**(6): p. 1166-8.
394. Schmitt, J., S. Langan, T. Stamm, H.C. Williams, et al., *Core outcome domains for controlled trials and clinical recordkeeping in eczema: international multiperspective Delphi consensus process.* Journal of Investigative Dermatology, 2011. **131**(3): p. 623-30.
395. Schmitt, J., P. Spuls, M. Boers, K. Thomas, et al., *Towards global consensus on outcome measures for atopic eczema research: results of the HOME II meeting.* Allergy, 2012. **67**(9): p. 1111-7.

396. Olliaro, P., M. Vaillant, B. Arana, M. Grogl, et al., *Methodology of clinical trials aimed at assessing interventions for cutaneous leishmaniasis.* PLoS Neglected Tropical Diseases [electronic resource], 2013. **7**(3): p. e2130.
397. Bellomo, R., C. Ronco, J.A. Kellum, R.L. Mehta, et al., *Acute renal failure - definition, outcome measures, animal models, fluid therapy and information technology needs: the Second International Consensus Conference of the Acute Dialysis Quality Initiative (ADQI) Group.* Critical Care (London, England), 2004. **8**(4): p. R204-12.
398. Molitoris, B.A., M.D. Okusa, P.M. Palevsky, P.L. Kimmel, et al., *Designing clinical trials in acute kidney injury.* Clinical Journal of the American Society of Nephrology, 2012. **7**(5): p. 842-843.
399. Endre, Z.H. and J.W. Pickering, *Acute kidney injury clinical trial design: old problems, new strategies.* Pediatric Nephrology, 2013. **28**(2): p. 207-17.
400. Abellan van Kan, G., W. Cameron Chumlea, S. Gillette-Guyonet, M. Houles, et al., *Clinical trials on sarcopenia: methodological issues regarding phase 3 trials.* Clinics in Geriatric Medicine, 2011. **27**(3): p. 471-82.
401. Devane, D., C.M. Begley, M. Clarke, D. Horey, et al., *Evaluating maternity care: a core set of outcome measures.* Birth, 2007. **34**(2): p. 164-72.
402. Bennett, W.L., K.A. Robinson, I.J. Saldanha, L.M. Wilson, et al., *High priority research needs for gestational diabetes mellitus.* Journal of Women's Health, 2012. **21**(9): p. 925-32.
403. Langguth, B., R. Goodey, A. Azevedo, A. Bjorne, et al., *Consensus for tinnitus patient assessment and treatment outcome measurement: Tinnitus Research Initiative meeting, Regensburg, July 2006.* Progress in Brain Research, 2007. **166**: p. 525-36.
404. Ramsey, B.W. and T.F. Boat, *Outcome measures for clinical trials in cystic fibrosis. Summary of a Cystic Fibrosis Foundation consensus conference.* Journal of Pediatrics, 1994. **124**(2): p. 177-92.
405. Gottrup, F., J. Apelqvist, P. Price, and G. European Wound Management Association Patient Outcome, *Outcomes in controlled and comparative studies on non-healing wounds: recommendations to improve the quality of evidence in wound management.* Journal of Wound Care, 2010. **19**(6): p. 237-68.
406. van Brussel, M., J. van der Net, E. Hulzebos, P.J.M. Helders, et al., *The Utrecht approach to exercise in chronic childhood conditions: the decade in review.* Pediatric Physical Therapy, 2011. **23**(1): p. 2-14.
407. Angus, D.C., J. Carlet, and P. Brussels Roundtable, *Surviving intensive care: a report from the 2002 Brussels Roundtable.* Intensive Care Medicine, 2003. **29**(3): p. 368-77.
408. Micke, O., M.H. Seegenschmiedt, and G. German Working Group on Radiotherapy in, *Consensus guidelines for radiation therapy of benign diseases: a multicenter approach in Germany.* International Journal of Radiation Oncology, Biology, Physics, 2002. **52**(2): p. 496-513.

Appendices

Appendix 1: Search strategies used in the systematic review of core outcome sets

	Search terms for MEDLINE	Number of hits (searched August 2013)
	Randomised trial and systematic review terms	
1	Health Services/ut [Utilization]	5742
2	registries/	45459
3	systematic review.mp.	28277
4	structured review.ti.	108
5	evidence based medicine.ab.	4843
6	exp Clinical Trials as Topic/	258536
7	clinical trial$.ab.	149831
8	randomised controlled trial$.ti,ab.	16321
9	randomised trial$.ti,ab.	9858
10	1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9	444510
	Methodology terms	
11	workgroup$.mp.	757
12	standard$ outcome$.mp.	519
13	Practice Guideline/	16919
14	clinical database.mp.	908
15	patient important outcome$.mp.	78
16	(standard$ adj3 reporting).mp.	1999

17	congresses.pt.	57907
18	Delphi Technique/	2274
19	(recommend$ adj3 outcome$).mp.	1062
20	consensus development conference.pt.	8085
21	outcome$ reporting.mp.	267
22	priorit$ symptom$.mp.	23
23	(task force adj3 outcome$).mp.	49
24	appropriate outcome$.mp.	338
25	research design/	67019
26	endpoint determination/	3416
27	consensus development conference/	8085
28	patient participation/	16143
29	consensus.mp.	93957
30	workshop.mp.	17486
31	Consensus Development Conferences, NIH as Topic/	314
32	focus groups/	13711
33	11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 or 31 or 32	279051
	Outcome terms	
34	outcome$.mp.	1156289
35	end point$.mp.	32706
36	(core adj3 set).mp.	1510

37	treatment emergent problem$.mp.	1
38	exp outcome Assessment Health Care/	593962
39	Treatment Outcome/	535004
40	Quality of Life/	101029
41	34 or 35 or 36 or 37 or 38 or 39 or 40	1256710
	Key terms targeted	
42	clinical-study design.mp.	82
43	patient$ perspective$.ti.	1387
44	outcome$.mp. and delphi.ti.	153
45	(outcome$ and delphi).ab.	624
46	(perspective$ adj3 outcome$).ti.	102
47	core outcome$.ti,ab.	121
48	core set$.ti,ab.	1124
49	clinical trial design$.ti.	355
50	design$ clinical trial$.ti.	72
51	(consensus and outcome$).ti.	133
52	42 or 43 or 44 or 45 or 46 or 47 or 48 or 49 or 50 or 51	3931
53	10 and 33 and 41	12607
54	52 or 53	16079
Search terms for SCOPUS		
((((INDEXTERMS(registries)) OR (INDEXTERMS(clinical trials as topic)) OR		12286

(ABS("evidence based medicine")) OR (ABS("clinical trial*")) OR (INDEXTERMS("Health Services Utilization")) OR (TITLE-ABS-KEY("SYSTEMATIC REVIEW")) OR (TITLE("structured review"))) OR (TITLE OR ABS("randomised controlled trial*")) OR (TITLE OR ABS (randomised trial*))) AND (((TITLE-ABS-KEY(workgroup*)) OR (TITLE-ABS-KEY(standard* outcome*)) OR (INDEXTERMS(practice guideline)) OR (TITLE-ABS-KEY("clinical database")) OR (TITLE-ABS-KEY("patient important outcome*")) OR (TITLE-ABS-KEY("standard* outcome*")) OR (INDEXTERMS(delphi technique))) OR ((TITLE-ABS-KEY(recommend* W/3 outcome*)) OR (TITLE-ABS-KEY(standard* W/3 reporting*)) OR (TITLE-ABS-KEY(task force W/3 outcome*)) OR (TITLE-ABS-KEY("appropriate outcome*")) OR (TITLE-ABS-KEY("outcome* reporting")) OR (TITLE-ABS-KEY("priorit* symptom*")) OR (INDEXTERMS(focus group)) (INDEXTERMS(research design))) OR ((INDEXTERMS(endpoint determination)) OR (INDEXTERMS(consensus development conference)) OR (INDEXTERMS(patient participation)) OR (TITLE-ABS-KEY(consensus)) OR (TITLE-ABS-KEY(workshop)))) AND 74) OR (((TITLE("design* clinical trials")) OR (TITLE(consensus AND outcome*)) OR (TITLE-ABS-KEY("clinical-study design")) OR (TITLE("patient* perspective*")) OR (ABS(outcome* AND delphi)) OR (TITLE(outcome* AND delphi)) OR (TITLE(perspective* W/3 outcome*)) OR (ABS("core outcome*") OR TITLE("core outcome*"))) OR ((ABS("core set*") OR TITLE("core set*")) OR (TITLE("clinical trial design*"))))	
Search terms for The Cochrane Library	

	Randomised trial and systematic review terms	
#1	(clinical trial*):ab	58990
#2	MeSH descriptor Health services	750
#3	MeSH descriptor registries	604
#4	(systematic review):ti,ab,kw	13816
#5	(structured review):ti	6890
#6	(evidence based medicine):ab	681
#7	MeSH descriptor Clinical Trials as Topic explode all trees	1117

#8	(randomised controlled trial):ti or (randomised controlled trial):ab	79593
#9	(randomised trial*):ti,ab,kw	158334
#10	(**#1** OR **#2** OR **#3** OR **#4** OR **#5** OR **#6** OR **#7** OR **#8 OR #9**)	126425
	Methodology terms	
#11	(workgroup*):ti,ab,kw	24
#12	MeSH descriptor Practice Guideline	1221
#13	(patient important outcome*):ti,ab,kw	3977
#14	(clinical database):ti,ab,kw	2335
#15	standard* NEAR/3 reporting	2310
#16	(congresses):pt	45
#17	MeSH descriptor Delphi Technique explode all trees	33
#18	recommend* NEAR/3 outcome	309
#19	(consensus development conference):pt	4
#20	(priorit* symptom*):ti,ab,kw	964
#21	(task force NEAR/3 outcome*):ti,ab,kw	4
#22	(appropriate outcome*):ti,ab,kw	2528
#23	MeSH descriptor Focus Groups explode all trees	232
#24	MeSH descriptor Research Design	1811
#25	MeSH descriptor endpoint determination	61
#26	MeSH descriptor consensus development conference	570

#27	MeSH descriptor patient participation	354
#28	(consensus):ti,ab,kw	2049
#29	(workshop):ti,ab,kw	955
#38	"standard outcome*":ti,ab,kw	27
#39	"outcome* reporting":ti,ab,kw	191
#37	(**#11 OR #12** OR **#13** OR **#14** OR **#15** OR **#16** OR **#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 **#38** OR **#39**)	14699
	Outcome terms	
#30	(outcome*):ti,ab,kw	143278
#31	(end point*):ti,ab,kw	12764
#32	(core NEAR/3 set):ti,ab,kw	82
#33	(treatment emergent problem*):ti,ab,kw	25
#34	MeSH descriptor Outcome Assessment (Health Care) explode all trees	81711
#35	MeSH descriptor Treatment Outcome	1957
#36	MeSH descriptor quality of life	1517
#40	(**#30** OR **#31** OR **#32** OR **#33** OR **#34** OR **#35** OR **#36**)	151831
#41	(#10 AND #37 AND #40)	7096
	Key terms targeted	
#42	(design* clinical trials):ti	541
#43	(clinical-study design):ti,ab,kw	2081
#44	(patient* perspective*):ti	161

#45	(outcome*):ti and (delphi):ti	0
#46	(outcome*):ab and (delphi):ab	62
#47	(perspective* NEAR/3 outcome*):ti	5
#48	(core outcome*):ti or (core outcome*):ab	616
#49	(core set):ti or (core set):ab	383
#50	(clinical trial design*):ti	541
#51	(outcome):ti and (consensus):ti	8
#52	(#42 OR #43 OR #44 OR #45 OR #46 OR #47 OR #48 OR #49 OR #50 OR #51)	3674
#53	(#41 OR #52)	10572
#54	(#53) *In Methods Studies*	1082

Appendix 2: Studies included in the systematic review (250 reports relating to 198 studies)

Study	Disease category	Disease name	Journal
Dixon 1987 [171]**	Cancer	Hodgkin's disease and lymphoma	Journal of Clinical Oncology
Glynne-Jones 2006 [172]**	Cancer	Rectal cancer	Annals of Oncology
Auvinen 1996 [173]**	Cancer	Prostate cancer	Journal of Medical Screening
Denis 1997 [174]*	Cancer	Prostate cancer (early stage)	Urology
Scher 2004 [175]*	Cancer	Prostate cancer (rising prostate-specific antigen)	Journal of Clinical Oncology
Dawson 1998 [176]*	Cancer	Prostate cancer	Journal of Clinical Oncology
Middleton 1995 [177] Schellhammer 1997 [178] **	Cancer	Localized prostate cancer	Journal of Urology Urology
Rajkumar 2011 [179]*	Cancer	Myeloma	Blood
Chow on behalf of International Bone Metastases Consensus Working Party [180]* Chow 2002 [181]	Cancer	Bone metastases	Clinical Oncology (Royal College of Radiologists) Radiotherapy & Oncology
Partsch 2010 [182]*	Cancer	Breast cancer related lymphedema (BCRL)	International Angiology
Hesketh 1998 [183]* Tonato 1998 [184]	Cancer	Chemotherapy-induced nausea and vomiting	Supportive Care in Cancer
Mcvie 1992 [185]*	Cancer	Chemotherapy-induced nausea and vomiting	Drugs
Pallis 2011 [186]**	Cancer	Solid tumours	Annals of Oncology
Miller 1981 [41]**	Cancer	Cancer (not specified)	Cancer
Prorok 2010 [187]*	Cancer	Cancer (not specified)	Seminars in Oncology
Punt [188]**	Cancer	Colorectal cancer	Journal of the National Cancer Institute
Wils 1998 [189]**	Cancer	Colorectal cancer (advanced)	Tumori
Llovet 2008 [190]*	Cancer	Hepatocellular carcinoma	Journal of the National Cancer Institute
Pagliusi 2004 [191]**	Cancer	Human papillomavirus (Cervical cancer)	Vaccine
Lefebvre 2009 [192]*	Cancer	Head and neck cancer	International Journal of Radiation Oncology, Biology, Physics
Adelstein 2012 [193]*	Cancer	Head and neck cancer	Head & Neck
Gridelli 2004 [194]*	Cancer	Advanced non-small-cell lung cancer	Annals of Oncology
Gridelli 2012 [195]*	Cancer	Advanced non-small-cell lung cancer	Lung Cancer
Cheson 1999 [196]** Cheson 2007 [197]	Cancer	Non-Hodgkin's lymphoma	Journal of Clinical Oncology

Study	Disease category	Disease name	Journal
Anderson 2008 [198]**	Cancer	Leukaemia	Leukemia
Cheson 2003 [199]*	Cancer	Acute Myeloid Leukaemia	Journal of Clinical Oncology
Kulke 2011 [200]*	Cancer	Neuroendocrine tumours	Journal of Clinical Oncology
Bellm 2002 [201]**	Cancer	Oral Mucositis (OM)	Cancer Investigation
Du Bois 2005 [202]* Stuart 2011 [203] Thigpen 2011 [204]	Cancer	Ovarian cancer	Annals of Oncology International Journal of Gynecological Cancer x2
Comenzo 2012 [205]*	Cancer	Systemic light-chain amyloidosis	Leukemia
Dorman 2009 [206]*	Cancer	Dyspnoea or Breathlessness in Palliative Care	Palliative Medicine
Renal Disease Subcommittee of the American College of Rheumatology Ad Hoc Committee on Systemic Lupus Erythematosus Response Criteria [207]*	Rheumatology	Systemic lupus erythematosus	Arthritis & Rheumatism
Bertsias 2009 [208]* Gordon 2009 [209]	Rheumatology	Systemic lupus erythematosus	Annals of the Rheumatic Diseases
Smolen 1999 [210]**	Rheumatology	Systemic lupus erythematosus	Journal of Rheumatology
Ruperto 2003 [211]**	Rheumatology	Juvenile systemic lupus erythematosus and juvenile dermatomyositis	Rheumatology
White 1995 [212]*	Rheumatology	Systemic sclerosis	Arthritis & Rheumatism
Khanna 2008 [213]**	Rheumatology	Systemic sclerosis	Annals of the Rheumatic Diseases
Clements 2012 [214]*	Rheumatology	Systemic sclerosis - related arthritis	Seminars in Arthritis and Rheumatism
Khanna 2010 [215]*	Rheumatology	Systemic sclerosis-associated interstitial lung disease	Clinical & Experimental Rheumatology
Merkel 2009 [216]** Merkel 2011 [217]	Rheumatology	Small-vessel Vasculitis/ ANCA-associated Vasculitis	Journal of Rheumatology
Hellmich 2007 [218]*	Rheumatology	Systemic vasculitis, AAV, anti-neutrophil cytoplasmic antibody-associated vasculitis; Wegener's granulomatosis	Annals of the Rheumatic Diseases
Felson 1993 [219]**	Rheumatology	Rheumatoid arthritis	Journal of Rheumatology
Fried 1993 [220]** Tugwell 1993 [221] Boers 1994 [42] Kirwan 2003 [45] Kirwan 2005 [44] Kirwan 2007 [46]	Rheumatology	Rheumatoid arthritis	Journal of Rheumatology
Sanderson 2010 [222]** Sanderson 2010 [223]	Rheumatology	Rheumatoid arthritis	Arthritis care & research

Study	Disease category	Disease name	Journal
Bombardier 1982 [224]**	Rheumatology	Rheumatoid arthritis	Journal of Rheumatology
Scott 1989 [225]**	Rheumatology	Rheumatoid arthritis	Annals of the Rheumatic Diseases
Van Riel 1992 [226]**	Rheumatology	Rheumatoid arthritis	British Journal of Rheumatology
Taylor 2005 [227]** Gladman 2005 [228] Gladman 2005 [229] Gladman 2007 [230]	Rheumatology	Psoriatic arthritis	Annals of the Rheumatic Diseases x3 Journal of Rheumatology
Miller 2001 [231]**	Rheumatology	Idiopathic inflammatory myopathies (IIM)	Rheumatology
Van Der Heijde 1997 [232]**	Rheumatology	Ankylosing spondylitis	Journal of Rheumatology
Wolfe 1999 [233]**	Rheumatology	Rheumatic diseases	Journal of Rheumatology
Mease 2005 [234]** Mease 2007 [235] Arnold 2008 [236] Mease 2008 [237] Carville 2008 [238] Choy 2009 [239] Mease 2009 [240]	Rheumatology	Fibromyalgia syndrome	Journal of Rheumatology x5 Patient Education & Counseling Arthritis & Rheumatism
Salaffi 2012 [241]**	Rheumatology	Fibromyalgia syndrome	Reumatismo
Schumacher 2005 [242]** Schumacher 2007 [243] Taylor 2008 [244] Schumacher 2009 [245]	Rheumatology	Gout	Journal of Rheumatology x 3 Annals of the Rheumatic Diseases
Giannini 1997 [246]**	Rheumatology	Arthritis	Arthritis & Rheumatism
Bellamy 1997 [247]**	Rheumatology	Knee, hip and hand osteoarthritis	Journal of Rheumatology
Heiligenhaus 2012 [248]**	Rheumatology	idiopathic arthritis-associated uveitis (juvenile)	Arthritis care & research
Bowman 2001 [249]** Pillemer 2005 [250]	Rheumatology	Sjögren's syndrome	Rheumatology Journal of Rheumatology
Cranney 1997 [81]** Guidelines for osteoporosis trials[251]	Rheumatology	Osteoarthritis	Journal of Rheumatology
Hammarlund 2012 [252]**	Neurology	Parkinson's disease	Quality of Life Research
Vellas 2008 [253]**	Neurology	Alzheimer's disease	Lancet Neurology
World Federation of Neurology Research Group 1995 [254]* Miller 1999 [255]	Neurology	Amyotrophic lateral sclerosis/motor neurone disease	Journal of the Neurological Sciences
Leigh 2004 [256]*	Neurology	Amyotrophic lateral sclerosis/motor neurone disease	Amyotrophic Lateral Sclerosis & Other Motor Neuron Disorders
Shankaran 2003 [257]*	Neurology	Hypoxic-ischemic brain injury	Seminars in Perinatology
Vargus-Adams 2009 [258]**	Neurology	Cerebral palsy	Archives of Physical Medicine & Rehabilitation

Study	Disease category	Disease name	Journal
Katona 2007 [259]**	Neurology	Dementia	International Psychogeriatrics
Moniz-Cook 2008 [260]**	Neurology	Dementia	Aging & Mental Health
ILAE Commission 1998 [261]*	Neurology	Epilepsy (newly diagnosed and chronic)	Epilepsia
LaFrance 2006 [262]*	Neurology	Seizures	Epilepsy & Behavior
Osborne 2001 [263]** Lux 2004 [264]	Neurology	Infantile spasms West syndrome (Epilepsy)	Brain & Development Epilepsia
Mindell 2006 [265]*	Neurology	Insomnia	Pediatrics
Schumacher 2010 [266]*	Neurology	Intracranial cerebral atherosclerosis	Journal of Neurointerventional Surgery
Whitaker 1995 [267]**	Neurology	Multiple sclerosis	Multiple Sclerosis
Chitnis 2012 [268]* Chitnis 2013 [269]	Neurology	Multiple sclerosis	Multiple Sclerosis Neurology
Penzien 2005 [270]* Penzien 2005 [271] Andrasik 2005 [272]	Neurology	Headache	Headache
Tfelt-Hansen 1991[273]* Tfelt-Hansen 2000 [274] Tfelt-Hansen 2012 [275]	Neurology	Migraine	Cephalalgia
Lipton 1995 [276]*	Neurology	Cluster headache	Cephalalgia
Schoenen 1995 [277]* Bendtsen 2010 [278]	Neurology	Tension-type headache	Cephalalgia
Hughes 2005 [279]*	Neurology	Chronic Inflammatory Demyelinating Polyradiculoneuropathy and Multifocal Motor Neuropathy	Neuromuscular Disorders
Merkies 2006 [280]**	Neurology	Peripheral neuropathy	Neuromuscular Disorders
Reilly 2006 [281]*	Neurology	Charcot-Marie-Tooth disease type 1A (CMT1A)	Neuromuscular Disorders
Clifton 1992 [282]**	Neurology	Traumatic brain injury	Neurosurgery
Wilde 2010 [283]**	Neurology	Traumatic brain injury	Archives of Physical Medicine & Rehabilitation
Duncan 2000 [284]**	Heart & circulation	Acute stroke	Stroke
Schellinger 2012 [285]**	Heart & circulation	Acute stroke	International Journal of Stroke
Hoeper 2004 [286]*	Heart & circulation	Pulmonary arterial hypertension	Journal of the American College of Cardiology
Distler 2008 [287]**	Heart & circulation	Pulmonary arterial hypertension relates Systemic Sclerosis	Arthritis & Rheumatism
Becker 2011 [288]**	Heart & circulation	Cardiac arrest	Circulation
Chiam 2008 [289]**	Heart & circulation	Aortic valve stenosis (AS)	Cardiovascular Interventions
Leon 2011 [290]** Kappetein 2012 [291]	Heart & circulation	Aortic stenosis (AS); Valvular heart disease	European Heart Journal European Journal of Cardio-Thoracic Surgery
Cutlip 2007 [292]**	Heart & circulation	Obstructive coronary artery disease	Circulation

Study	Disease category	Disease name	Journal
Simons 2000 [293]*	Heart & circulation	Ischemic heart disease	Circulation
Conte 2009 [294]**	Heart & circulation	Critical limb ischemia	Journal of Vascular Surgery
Timaran 2011 [295]*	Heart & circulation	Atherosclerosis	Journal of Vascular Surgery
Stout 2012 [296]*	Heart & circulation	Chronic leg edema	International Angiology
Nedeltchev 2010 [297]**	Heart & circulation	Obstructive disease of supra-aortic arteries	Catheterization & Cardiovascular Interventions
Anderson 2013 [298]*	Heart & circulation	Cardiovascular disease	Journal of the American College of Cardiology
Hausenloy 2013 [299]*	Heart & circulation	Coronary heart disease	Cardiovascular Research
Labs 1999 [300]*	Heart & circulation	Peripheral arterial occlusive disease (PAOD)	European Journal of Vascular & Endovascular Surgery
Mitchell 2011 [301]**	Heart & circulation	Deep venous thrombosis and pulmonary embolism	Journal of Thrombosis & Haemostasis
Steg 2011 [302]*	Heart & circulation	Acute coronary syndrome	European Heart Journal
O'Connell 2009 [303]*	Heart & circulation	Acute heart failure syndromes (AKA Acute decompensated heart failure)	Heart Failure Reviews
Higashida 2003 [304]*	Heart & circulation	Acute ischemic stroke	Stroke
Calkins 2012 [305]*	Heart & circulation	Atrial fibrillation	Journal of Interventional Cardiac Electrophysiology
Kirchhof 2007 [306]**	Heart & circulation	Atrial fibrillation	European Heart Journal
Buser 1997 [307]*	Dentistry & oral health	Implants in regenerated bone	Annals of Periodontology
Page 1992 [308]*	Dentistry & oral health	Periodontitis	Journal of Periodontal Research
Imrey 1994 [309]*	Dentistry & oral health	Periodontitis	Journal of Periodontal Research
Lightfoot 2005 [310]**	Dentistry & oral health	Chronic periodontitis (Anterior teeth)	Journal of Periodontology
Lightfoot 2005 [311]**	Dentistry & oral health	Chronic periodontitis (Posterior teeth)	Journal of Periodontology
Weber 1997 [312]*	Dentistry & oral health	Endentulous	Annals of Periodontology
Cochran 1998 [313]*	Dentistry & oral health	Endentulous	Journal of Periodontology
Tonetti 2012 [314]**	Dentistry & oral health	Endentulous	Journal of Clinical Periodontology
Smaïl-Faugerson 2013 [315]**	Dentistry & oral health	Extensive tooth decay	PLoS ONE
Chilton 1986 [316]*	Dentistry & oral health	Plaque and gingivitis	Journal of Clinical Periodontology
Council on Dental Therapeutics 1986 [317]*	Dentistry & oral health	Supragingival dental plaque and gingivitis	Journal of the American Dental Association
Pitts 2004 [318]*	Dentistry & oral health	Caries	Journal of Dental Research
Marshall 2005 [319]**	Infectious disease	Sepsis and critical care	Critical Care Medicine
Goldstein 2005 [320]*	Infectious disease	Sepsis and critical care	Pediatric Critical Care Medicine

Study	Disease category	Disease name	Journal
Wood 1995 [321]*	Infectious disease	Herpes Zoster	Journal of Antimicrobial Chemotherapy
Barlow 2003 [322]**	Infectious disease	Community-acquired pneumonia	The Lancet Infectious Diseases
Spellberg 2008 [323]*	Infectious disease	Community-acquired pneumonia	Clinical Infectious Diseases
Powers 2010 [324]* Spellberg 2010 [325]	Infectious disease	Hospital-acquired pneumonia and ventilator-associated pneumonia	Clinical Infectious Diseases
McCracken 1992 [326]*	Infectious disease	Acute bacterial meningitis	Clinical Infectious Diseases
Alioum 2001 [327]*	Infectious disease	HIV	Statistics in Medicine
Kirkby 2010 [328]*	Infectious disease	Influenza	Journal of Alternative & Complementary Medicine
Nystrom 1990 [329]*	Infectious disease	Intraabdominal infection	World Journal of Surgery
Cross 2005 [330]**	Infectious disease	Leprosy	International Journal of Leprosy & Other Mycobacterial Diseases
Moorthy 2007 [331]** Moorthy 2009 [332]	Infectious disease	Malaria	Vaccine
Steeves 2007 [333]**	Orthopaedics & trauma	Spinal cord injury	Spinal Cord
Bombardier 2000 [334]**	Orthopaedics & trauma	Spinal disorders	Spine
Deyo 1998 [335]**	Orthopaedics & trauma	Low back pain	Spine
Devogelaer 2003 [336]*	Orthopaedics & trauma	Acute low back pain	Clinical & Experimental Rheumatology
Goldhahn 2008 [337]**	Orthopaedics & trauma	Osteoporosis	Bone
Lynch 2013 [338]**	Orthopaedics & trauma	ACL injury	British Journal of Sports Medicine
Falder 2009 [339]**	Orthopaedics & trauma	Burns	Burns
Reneman 2013 [340]**	Orthopaedics & trauma	Musculoskeletal pain (subacute and chronic)	Journal of Occupational Rehabilitation
Lamb 2005 [341]**	Orthopaedics & trauma	Fall injury	Journal of the American Geriatrics Society
Cameron 2010 [342]*	Orthopaedics & trauma	Hip fracture	Osteoporosis International
Smith 1996 [343]**	Lungs & airways	Asthma	Australian & New Zealand Journal of Public Health
Reddel 2009 [104]**	Lungs & airways	Asthma	American Journal of Respiratory Crital Care Medicine
Busse 2012 [105]** Fuhlbrigge 2012 [344] Akinbami 2012 [345] Szefler 2012 [346] Tepper 2012 [347] Cloutier 2012 [348] Krishnan 2012 [349] Wilson 2012 [350]	Lungs & airways	Asthma	Journal of Allergy & Clinical Immunology
Sinha 2012 [103]**	Lungs & airways	Asthma	Trials

Study	Disease category	Disease name	Journal
Keim 2004 [351]**	Lungs & airways	Respiratory distress	Academic Emergence Medicine
Canonica 2007 [352]*	Lungs & airways	Respiratory allergy	Allergy
Cazzola 2008 [353]**	Lungs & airways	Chronic obstructive pulmonary disorder (COPD)	European Respiratory Journal
Task Group on Mucoactive Drugs 1994 [354]*	Lungs & airways	Chronic bronchitis and COPD	Chest
Dent 2008 [355]*	Gastroenterology	Gastro-oesophageal reflux disease (GERD)	Alimentary Pharmacology & Therapeutics
Wirth 2011 [356]*	Gastroenterology	Chronic Hepatitis C	Journal of Pediatric Gastroenterology & Nutrition
Bajaj 2011 [357]*	Gastroenterology	Hepatic encephalopathy	Alimentary Pharmacology & Therapeutics
Fekety 1992 [358]*	Gastroenterology	Antibiotic associated colitis	Clinical Infectious Diseases
Griffiths 2005 [359]**	Gastroenterology	Crohn's disease	Inflammatory Bowel Diseases
Laine 2010 [360]*	Gastroenterology	Nonvariceal upper gastrointestinal bleeding	American Journal of Gastroenterology
Pimentel 2013 [361]*	Gastroenterology	Irritable bowel syndrome (IBS)	Gastroenterology
Sanyal 2011 [362]*	Gastroenterology	Nonalcoholic steatohepatitis	Hepatology
Rahn 2011 [363]**	Gynaecology	Abnormal uterine bleeding	Journal of Clinical Epidemiology
Meuleman 2012 [364]**	Gynaecology	Deeply infiltrative endometriosis (DIE)	Current Opinion in Obstetrics & Gynecology
Vincent 2010 [365]*	Gynaecology	Endometriosis-related pain	Fertility & Sterility
Broder 2000 [366]*	Gynaecology	Uterine fibroids	Journal of Vascular & Interventional Radiology
Basson 2000 [367]*	Gynaecology	Female sexual dysfunction	Journal of Urology
Clayton 2010 [368]*	Gynaecology	Female sexual dysfunction	Journal of Sexual Medicine
Walker 2006 [369]**	Tobacco, drugs, & alcohol dependence	Addiction (gambling)	Addiction
Del Boca 2007 [370]*	Tobacco, drugs, & alcohol dependence	Addiction (substance)	Addiction
Donovan 2012 [371]**	Tobacco, drugs, & alcohol dependence	Drug dependence	Addiction
Vocci 1999 [372]*	Tobacco, drugs, & alcohol dependence	Nicotine, alcohol and cocaine abuse/dependence	National Institute on Drug Abuse Medications Development Division
Levine 2003 [373]**	Urology	Peyronie's disease	International Journal of Impotence Research
Djurhuus 1997 [374]**	Urology	Nocturnal enuresis	Scandinavian Journal of Urology & Nephrology

Study	Disease category	Disease name	Journal
Toozs-Hobson 2012 [375]**	Urology	Pelvic organ prolapse	International Urogynecology Journal
Porst 2010 [376]*	Urology	Male sexual dysfunction/disorders	Journal of Sexual Medicine
Pavletic 2006 [377]**	Blood disorders	Chronic graft-versus-host disease (GVHD)	Biology of Blood & Marrow Transplantation
Lassila 2005 [378]*	Blood disorders	Haemophilia and other bleeding disorders	Haemophilia
Rodeghiero 2009 [379]**	Blood disorders	Immune thrombocytopenic purpura	Blood
Turk 2003 [125]** Turk 2008 [380]	Anaesthesia & pain control	Chronic pain	Pain
McGrath 2008 [124]**	Anaesthesia & pain control	Chronic pain/ recurrent pain and acute pain	Journal of Pain
Apfel 2002 [381]*	Anaesthesia & pain control	Post-operative nausea and vomiting	Anaesth Intensivmed Notfallmed Schmerzther
Anderson 1998 [382]*	Endocrine & metabolic	Obesity	Obesity Research
Douglas 2009 [383]**	Endocrine & metabolic	Thyroid eye disease (TED)	Archives of Ophthalmology
Carlson 2003 [384]*	Mental health	Bipolar disorder	Journal of Child & Adolescent Psychopharmacology
Rush 2006 [385]**	Mental health	Major depressive disorder	Neuropsychopharmacology
Fitzpatrick 2010 [386]**	Mental health	Forensic mental health	Health Technology Assessment
Finer 2006 [387]* Giacoia 2006 [80]	Neonatal care	Neonatal apnea (also known as Apnea of prematurity, and Apnoea)	Pediatrics
Short 2006 [388]* Giacoia 2006	Neonatal care	Neonatal cardiovascular instability	Pediatrics
Clancy 2006 [389]* Giacoia 2006	Neonatal care	Neonatal seizures	Pediatrics
Gonzalez 2011 [390] Eleftheriadou 2012 [391]*	Skin	Vitiligo	Archives of Dermatology British Journal of Dermatology
Schmitt 2007 [392]** Schmitt 2010 [393] Schmitt 2011 [394] Schmitt 2012 [395]	Skin	Eczema	Journal of Allergy & Clinical Immunology British Journal of Dermatology Journal of Investigative Dermatology Allergy
Olliaro 2013 [396]*	Skin	Cutaneous leishmaniasis	PLoS Neglected Tropical Diseases
Bellomo 2004 [397]*	Kidney disease	Acute renal failure	Critical Care
Molitoris 2012 [398]*	Kidney disease	Acute kidney injury	Clinical Journal of the American Society of Nephrology
Endre 2013 [399]*	Kidney disease	Acute kidney injury	Pediatric Nephrology

Study	Disease category	Disease name	Journal
Abellan van Kan 2011 [400]*	Health care of older people	Sarcopenia	Clinics in Geriatric Medicine
Devane 2007 [401]**	Pregnancy & child birth	Maternity care	Birth
Bennett 2012 [402]*	Pregnancy & child birth	Gestational diabetes mellitus	Journal of Women's Health
Langguth 2007 [403]**	Ear, nose & throat	Tinnitus	Progress in Brain Research
Ramsey 1994 [404]**	Genetic disorders	Cystic Fibrosis	Journal of Pediatrics
Gottrup 2010 [405]**	Wounds	Non-healing wounds	Journal of Wound Care
van Brussel 2011 [406]*	Chronic conditions	JIA, OI, Achondroplasia, Hemophilia, Cerebral, Palsy, Spina Bifida, CF, Cancer	Pediatric Physical Therapy
Angus 2003 [407]*	Intensive care	Critical illness/ ICU disease	Intensive Care Medicine
Micke 2002 [408]*	Benign disease	Benign/ Non-malignant diseases	International Journal of Radiation Oncology, Biology, Physics

** Considered outcomes while addressing wider clinical trial design issues*
*** Specifically considered outcome selection and measurement*

Appendix 3: A description of rounds in studies that used the Delphi technique

Study	Number of rounds	Round 1	Round 2	Round 3	Other rounds	How outcomes were kept in between rounds
White (1995)	Unclear	Number of rounds unclear				Unclear
Broder (2000)	2	Outcomes were rated independently and anonymously (either “important to measure” or “essential to measure” or "do not measure").	The outcomes with the highest group rating were discussed and several new ones added. Independent rating took place.	N/A	N/A	Round 2: The outcomes with the highest group rating were discussed and several new ones added and outcomes were again rated the second time. 41 outcomes got reduced to 23 outcomes
Basson (2000)	2	Unclear	Unclear	N/A	N/A	Unclear
Lightfoot (2005)	2	Rate outcome measures as either “extremely important,” “very important,” “moderately important,” “somewhat important,” or “not important” for successful therapy in relation to the scenario.	Requested to review the most frequent responses to the first round of the survey and if they agreed to affirm the result or if they disagreed, to mark what they considered the most appropriate outcome.	N/A	N/A	Round 2: The most frequent responses of Round 1 were reviewed.
Lightfoot (2005)	2	Rate outcome measures as either “extremely important,” “very important,” “moderately important,” “somewhat important,” or “not important” for successful therapy in relation to the scenario.	Requested to review the most frequent responses to the first round of the survey and if they agreed to affirm the result or if they disagreed, to mark what they considered the most appropriate outcome.	N/A	N/A	Round 2: The most frequent responses of Round 1 were reviewed.

Sinha(2012)	2	Open questions were asked in order to identify a long list of outcomes.	To identify the relative importance of each outcome, participants were asked the following question: "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". They were also asked to pick the three outcomes they felt were most important. In order to ensure that important outcomes were not missed, participants were asked to suggest any unlisted outcomes that they would have selected in their top 3.	N/A	N/A	Round 2: To enable each group of participants, regardless of its size, to have equal opportunity to suggest outcomes for phase 2, those outcomes suggested by at least 10 % of young people and/or parents and/or clinicians were carried forward to the next phase. By censoring in this way, we reduced the number of outcomes listed on the phase 2 questionnaire, without overlooking outcomes of potentially genuine importance. The reviewers discussed the individual outcomes that had not been suggested by sufficient numbers of participants, but were measured in at least 10 % of RCTs identified in the systematic review described earlier. If nearly 10 % of both clinicians and parents suggested the outcome, it was carried forward to the next phase, because we felt that, if we had a larger sample size, the outcome may have been suggested by sufficient numbers of

						participants.
Schmitt (2011)	3	Rate outcome domains on a nine-point Likert scale (1-3=not important, 4-6=Equivocal, 7-9=important) in the context of (a) clinical trials and (b) recordkeeping in daily practice. Asked to list additional outcome domains they considered potentially relevant. Additionally, they were asked in the first round to indicate how many domains should be included in the final core set of outcome domains for each context.	In subsequent rounds, participants received feedback on their own response along with the group opinion for each domain (median and interquartile range, calculated using Stata 10, Stata, College Station, TX) from the previous round. Respondents could submit new scores or leave their scores unchanged.	In the final round, instead of ranking the importance of the individual domains on a Likert scale, participants were asked explicitly which domains they recommend incorporating into the core set.	N/A	Round 3: Although the panel considered a broad set of different outcome domains as important, the panel indicated that only three different domains should be included in the core set (median rating by the whole panel).

Vargus-Adams (2009)	3	Open question asked in order to identify a long list of outcomes.	The second survey listed the domains that were identified from the first survey and asked the respondents to rank them in order of importance. Respondents were asked to rank every item, starting with 1 for the item they thought was most important.	The third survey again addressed these domains and asked the respondents to demonstrate the relative value of each domain by distributing 100 points among the domains. The ranks for each domain were examined for any patterns or obvious drop-off points that would permit the elimination of one or more domains that were less valued by the respondents. This process led to the final list of important domains.	Round 4 and 5 related to how to measure those outcomes.	After round 3, the ranks for each domain were examined for any patterns or obvious drop-off points that would permit the elimination of one or more domains that were less valued by the respondents.

Cross (2005)	4	Open question asked in order to identify a long list of criteria.	Rating/ranking of collated criteria.	Round 3: Further refinement of agreed criteria. Participants were required to again consider the criteria and given a concise description of outcomes.	Round 4: On receipt of the response from Round 3, outcomes were tabulated as 'Draft gold standard'. Assurance was given that should 4 or more people request changes to any one item in the tables, such request would be implemented. It was re-iterated that screening, assessment and outcome criteria were no longer negotiable.	Each rank was given a score. Descending negative scores were given when members had ranked a criterion as either 'not useful' or 'should be omitted'. If a criterion was ranked as 'neutral' it was scored as O. Ascending positive scores were given where members had ranked a criterion as either 'useful' or 'essential'. With nine members contributing the final score for each criterion represented the mean of the nine responses. Score>1. The indication was the criterion is not acceptable and should be rejected. Score >1<2 acceptable and should therefore be included for further consideration.
Heiligenhaus (2012)	Unclear	Number of rounds unclear. Each domain and item was ranked on a scale from 1–5, where 1 was of highest importance and 5 lowest importance.				Domains or items with low importance or redundant variables were eliminated.

Salaffi (2012)	1	Clinicians rated symptoms/domains using a Likert scale from 1 to 4 (4=highly relevant, extremely important; 3=very relevant, very important; 2=not very relevant, not very important; 1=not relevant, unimportant). Patients asked to put various domains in order of priority giving each one a score on the Likert scale from 1 to 3.	N/A	N/A	N/A	In order to reduce the number of clinical domains, items were excluded from the list if: a) they were related to gender; b) they required use of special equipment; c) they used terminology which was ambiguous or difficult to understand; d) they presented alternatives to other items, duplicated them and/or were similar to them.
Smaïl-Faugeron(2013)	3	Rate the importance of each outcome on a 5-point Likert-type scale: 1, no importance; 2, some importance; 3, moderate importance; 4, very important; and 5, extremely important.	The results from the first round were relayed back to participants. Participants were asked to choose the outcomes that should be part of the core set.	Round 3 sought to obtain broader consensus on the core set of outcomes. The results from round 2 were relayed back to participants. The participants were asked to re-select the outcomes that should be part of the core set with	N/A	Round 2: Only outcomes or component outcomes rated very or extremely important by at least 50% of participants were carried forward to round 2. Round 3: Component outcome measures chosen by 70% or more of participants were retained in round 3.

				knowledge of the group's previous ratings. Outcome measures chosen by 70% or more of participants were retained in round 3.		
Bennett (2012)	3 [only round 3 for voting]	Outcomes suggested.	Outcomes suggested.	Round 3 for outcomes prioritisation: From a list of possible outcomes that had been suggested in rounds 1-2. Each stakeholder ranked their top three outcomes that would be most important to include in a clinical trial that assessed medication and delivery	N/A	N/A. The aim was to prioritise, not reach consensus on the final set

				management.		
Rahn (2011)	Unclear	Number of rounds unclear. The importance of each outcome was graded on a scale suggested by the GRADE working group: a 9-point scale with outcomes scored as ‘‘critical’’ for decision making (score: 7-9), ‘‘important but not critical’’ (score: 4-6), and ‘‘not important’’ for decision making (score: 1-3).				During outcome categorisation: From the outcome inventory, the outcomes were organized and grouped into eight proposed overarching outcome domains. Outcomes related to cost, resource use, or those determined by the review group to have limited relevance for assessing clinical effectiveness were excluded from categorization and further analyses
Dent (2008)	3	Unclear (between each of the three voting rounds, statements were revised based on feedback from the Working Group and additional literature reviews, and some statements were added on matters not addressed previously)	Unclear	Unclear	N/A	Unclear

Devane (2007)	3	Rate the importance of each outcome listed using a 5-point Likert-type scale rating their importance for inclusion in a minimum set as: 1 = of no importance, 2 = of some importance, 3 = of moderate importance, 4 = very important, and 5 = extremely important. Participants were also asked to identify up to 2 ‘‘new’’ outcomes under each of 5 broad headings, which they judged to be relevant or important.	Participants were asked to re-rate the importance of each outcome with knowledge of their individual and the group’s previous ratings. In addition, participants were asked to rate the newly identified outcomes from round 1. All ratings used the same Likert-type scale that was used in round 1.	Each of the outcomes in round 3 was again presented together with the mean rating and standard deviation for the whole group, and participants were asked to re-rate the importance of each item for inclusion in a minimum data set using the same Likert-type scale used in round 2.	N/A	Round 2: outcomes retained after analysis of responses from round 1, where (a) the overall mean score for inclusion for that outcome was greater than the mean score for all the outcomes combined and (b) the mean score for inclusion for that outcome was greater than the mean score for all the outcomes combined for those participants who had rated their perceived level of expertise in evaluating models of maternity care as high (i.e., 6 or 7 on the Likert scale). Round 3: Outcomes retained after analysis of responses from round 2, where (a) the overall mean score for inclusion for that outcome was greater than the mean score for all the outcomes combined and (b) 70 percent or more of study participants rated their importance for inclusion as a ‘‘4’’ or ‘‘5’’ on the 5-point Likert-type scale used in round 2.

Distler (2008)	3	Score each domain and tool for use as outcome measures in randomized controlled trials using a 5-point scale, where a score of 1 indicated “not important/appropriate at all” and 5 indicated “very important/appropriate.” A text box of unlimited size was provided for free text below each domain and its associated measurement tools to add new tools. Additional domains could be proposed at the end of the questionnaire.	Participants were asked to repeat the rating of the domains and tools based on the information from the group rating of stage 1.	Participants were asked to perform another, and final, rating of the items retained from round 2. As in stage 2, participants were shown their own rating in the previous stage as well as the median ratings of the entire group.	N/A	A cluster analysis was performed on the items from stages 2 and 3 to differentiate important/appropriate from unimportant/inappropriate domains and tools. This reduced the number of domains and tools in a statistically significant manner.
Douglas (2009)	3	To identify parameters that they believed could be used in a 1-year multicentre clinical trial.	To rate the importance of each item. Each respondent rated the criteria on a scale of 1 (extremely inappropriate for a combined measure) to 9 (extremely appropriate for a combined measure)	A report containing the final questionnaire was sent to participants that provided feedback to the respondents, reminding them of their previous ratings in Delphi 2 for each criterion	N/A	Round 3: Steering committee review and RAND/University of California, Los Angeles, appropriateness method.

				compared with a group mean (standard deviation). The questionnaire requested that each participant again rate the criteria after they considered the mean group response.		
Khanna (2008)	3	List the items in the 11 domains suggested that could be used in development of a response index for a 1-year multi-centre clinical trial.	They were asked to rate each item on a scale of 1 (extremely inappropriate for a combined measure) to 9 (extremely appropriate for a combined measure).	Round 3 [after NGT meeting] they were asked to provide their agreement on a scale of 1 (complete disagreement on the core set measure for a combined measure) to 9 (complete agreement on the core set measure for a combined measure).	N/A	Round 3: Steering committee review and RAND Corp./University of California, Los Angeles (RAND/UCLA) Appropriateness Method

Lux (2004)	6	Outcomes suggested.	Outcomes suggested.	Participants were invited to respond to 32 statements that represented majority opinions from earlier rounds and to state whether they agree, unsure or would prefer an amendment. They also were asked to state whether they thought the statement should be included in the final proposal.	Round 4: presented modified statements requesting comments on their content and suitability for the proposal. Round 5: consisted of a draft paper for submission as a final proposal from the West Delphi Group. Round 6: consisted of a final approval of the draft paper.	Unclear

Mease (2005, 2008)	Clinicians: 3 Patient: 2	Clinician Delphi: Participants asked to distribute 100 points among the domains, giving more points to domains they considered more important to evaluate. Patient Delphi: participants were asked to rank each domain as it applied to them and impacted their life. Participants were instructed to distribute 100 points among the domains.	Clinician Delphi: Result of previous round group median, interquartile range, and total range of the earlier responses were provided to reflect on them and revise subsequent scoring if they choose. Patients Delphi: Participants were asked to rank each domain as it applied to them and impacted their life. Participants were presented with their score on each domain from the previous round and the domain ranking results from round 1 for all patients, which were revealed as the mean and total range.	Clinician Delphi: Result of previous round group median, interquartile range, and total range of the earlier responses were provided to reflect on them and revise subsequent scoring if they choose.	N/A	Delphi was conducted to establish a prioritized list of key domains before a workshop.
McGrath (2008)	2	Rate the outcome domains that had been suggested in original recommendations and to recommend others. They were also asked to suggest possible measures for each domain.	Results of the first round were summarized and were supplied to participants for the repeat poll.	N/A	N/A	Unclear
Miller (2001)	2	Unclear	Unclear	N/A	N/A	Unclear

Ruperto (2003)	2	Indicate up to 10 variables that they judged as clinically most important. Any types of variable (e.g. laboratory tests, questionnaires, indices of disease activity) could be chosen.	Variables indicated by at least 10 responders in the first round and additional variables used in published therapeutic trials or observational studies were listed in alphabetical order. Among these variables, physicians were asked to select, in order of importance, their top 10 choices. They were also asked to define the minimum and maximum number of variables that should be included in the core set.	N/A	N/A	Round 2: Variables indicated by at least 10 responders in the first round and additional variables used in published in therapeutic trials or observational studies were listed in Round 2.
Taylor (2005)	3	Distribute 100 points amongst the domains. Respondents were free to assign points to as many or as few domains as they wished.	Participants were given their own response and the group opinion for each domain (median and interquartile range) from the previous round. Respondents were able to submit new scores or leave their scores unchanged.	Participants were given their own response and the group opinion for each domain (median and interquartile range) from the previous round. Respondents were able to submit new scores or leave their scores unchanged.	N/A	Using structured consensus methods, this exercise reduced the number of potential domains for consideration in a core set from 26 to around a dozen before the consensus meetings. Outcomes were not removed rather divided into higher and lower scoring domains.

Taylor (2008)	3	Rate the importance of domains on a seven-point scale (1 = definitely necessary to 7 = definitely not necessary). Additional domains felt to be of importance could be added.	New items, re-worded items, and items for which there was disagreement and/or median rating of 4 (neither agreement nor disagreement) were re-rated in the second iteration.	In the final (third) iteration, no new items were introduced and only items for which there was disagreement and/or median rating of 4 were re-rated.	N/A	Rating of 5-7=should not be included

Appendix 4: Invite to take part in the COS developer survey

From: Gargon, Liz

Sent: 08 October 2014 11:57

To:

Subject: Re your study: [Study title]

Dear [Name]

We have recently completed a systematic review to identify studies that sought to determine which outcomes or domains to measure in all clinical trials in a specific condition. Your study [Title] was identified in our search and included in the review. You can view the systematic review at

http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0099111

As part of a follow up study from this review we are writing to ask if you'd be willing to answer a few short questions about your study. We have prepared a short questionnaire to help with this and would be very grateful if you'd be willing to fill this in. Our follow up study will contribute to a larger research project that is aiming to develop methodological guidance for core outcome set development. In order to develop this guidance we need to try to understand what factors have informed the ways in which researchers have developed core outcome sets. We would like to learn from your experience of doing this and incorporate what we learn from you into guidance to assist future researchers. We will acknowledge your contribution in the publication of the results, but all responses will remain confidential and data will be aggregated.

The questionnaire can be accessed online Survey.liv.ac.uk/version1 and should take around 15 minutes to complete. Your progress can be saved and returned to at any point via the 'save' button at the bottom of each page. We ask that you please complete the questionnaire within two weeks of receiving this email.

Thank you very much for taking the time to consider this request.

With best wishes

Liz Gargon

Appendix 5: COS developer survey

Page 1

Welcome to the Core Outcome Sets questionnaire

The COMET (Core Outcome Measures in Effectiveness Trials) Initiative (www.comet-initiative.org) is bringing together people interested in the development and application of core outcome sets (COS). COMET aims to collate and stimulate relevant resources, both applied and methodological, to facilitate exchange of ideas and information, and to foster methodological research in the area of COS. We have recently completed a systematic review to identify studies which sought to determine which outcomes or domains to measure in all clinical trials in a specific condition. Using this approach we have identified 198 studies. Your study was identified in our search and we have included it in the review and the COMET database.

We would like to learn from real life experiences and incorporate what we learn into guidance to assist future researchers. As such, I would like to ask you to answer a few short questions about your work. The survey will take around 15 minutes to complete. You can save the questionnaire and return to it at any point.

Please click 'next' to complete the questionnaire.

Thank you
Liz Gargon, COMET Project Coordinator

Page 2

Your details

1. **Please enter your name:**

2. **Please enter your email address:**

Page 3

Core outcome set details

With regard to the core outcome set (COS) study that we referred to in the email we sent you, please clarify the below details about your COS.

3. This core outcome set is intended to be applied to studies involving: [dropdown]

Both adults and children
Adults
Children (including neonates)
Older adults only

4. This core outcome set is intended to be applied to studies involving the following interventions: [check list –tick all that apply]

Devices
Procedures
Surgery
Vaccines
Drug treatments
Behavioural
All intervention types
Other, please specify [free text to specify]

5. Please indicate which methods you used to develop the core outcome set [check list –tick all that apply]

- ☐ Literature/systematic review
- ☐ Nominal Group Technique
- ☐ Non Delphi survey (e.g. in a questionnaire)
- ☐ Delphi (a structured technique to reach consensus)
- ☐ Meeting (an assembly of people for a particular purpose, particularly for discussion)
- ☐ Consensus development conference (key features include summaries of current knowledge, audience participation and a panel to assess the evidence presented
- ☐ Other, please specify

6. Please indicate which participants you included in the development of this core outcome set [check list –tick all that apply]

- ☐ Clinical professionals (e.g. physicians, nurses, physiotherapists, counsellors, occupational therapists)
- ☐ Patients
- ☐ Caregivers (informal e.g. family, friends)
- ☐ Patient/support group representatives
- ☐ Researchers - non clinical
- ☐ Researchers - clinical
- ☐ Industry representatives
- ☐ Charity representatives
- ☐ Governmental agencies (e.g. NIH)
- ☐ Policy makers (e.g. technology assessment and health policy)
- ☐ Regulatory agency representatives (e.g. FDA, EMA)
- ☐ Other, please specify

Page 4

Methods

7. **How did your core outcome set study come about?**

   Why did you/your group decide to develop the core outcome set? (Please tick all that apply)

- ☐ It was part of a research prioritisation study*
- ☐ I/we thought there was something missing in the outcomes being measured/reported in research
- ☐ I/we thought there was something missing in the outcomes being measured/reported in clinical practice
- ☐ The outcomes that were being measured in research were not applicable/relevant to clinical practice
- ☐ I/we were motivated by work that had been done in another speciality
- ☐ There was no existing core outcome set that we could use for our study
- ☐ A core outcome set existed but we did not think it was good enough/suitable*
- ☐ There was heterogeneity in which outcomes were being measured (studies measuring/reporting different outcomes) in trials/research
- ☐ There was heterogeneity in the way outcomes were being measured (studies measuring the same outcomes using different tools/instruments/measures) in trials/research
- ☐ There were outcomes being measured but not reported in trials/research
- ☐ I/we were conducting other research (e.g. a systematic review or a trial)*
- ☐ Other, please specify

8. **How did you decide on the methods you used?**

   (Please tick all that apply)

- ☐ Based on the literature (previous work)*
- ☐ Problems with other methods*

- ☐ Suited our situation and circumstances*
- ☐ Based on the resources available*
- ☐ Based on expert advice*
- ☐ Own experience with same methods before for core outcome set development
- ☐ Other, please specify

A conditional question is set up, so that when each answer marked with a * is ticked, an additional box appears below asking to provide more information about those responses.

9. **How did you decide who to include as participants in your core outcome set development work?**
   (Please tick all that apply)

- ☐ Experience with/knowledgeable about clinical practice
- ☐ Experience with/knowledgeable about trials/research
- ☐ Experience of living with/having/caring for someone with the condition
- ☐ Able to see things from the patient perspective
- ☐ Experience of people who are involved in decision-making about treatment
- ☐ To help with implementation and uptake later on
- ☐ To represent a broad view
- ☐ We wanted a local perspective
- ☐ We wanted a national perspective
- ☐ We wanted an international perspective
- ☐ Other, please specify

Page 5

Methods

10. **Were there any differences in the outcomes thought to be important by different stakeholder groups?** [dropdown]

Yes
No
We did not look at this
N/A – only one group of stakeholders included

A conditional question is set up, so if yes is selected, another box appears below asking for more information.

Page 6

Your experiences

11. What do you think were the main <u>strengths</u> of your study? *(Please list up to 5)*

1)
2)
3)
4)
5)

12. What do you think were the main <u>challenges</u> you experienced over the course of your study? *(Please list up to 5)*

1)
2)
3)
4)
5)

13. What do you think were the main <u>limitations</u> of your study? *(Please list up to 5)*

1)
2)

3)

4)

5)

14. Reflecting on your experiences of developing a core outcome set, are there any areas that you feel would benefit from methodological guidance or research to inform future activity to develop core outcome sets? *(Please list up to 5)*

1)

2)

3)

4)

5)

Page 7

Resources

15. Can you please list the resources you used to develop your core outcome set?
Resources available (e.g. funding for salary of researcher, travel expenses for meeting participants)

FREE TEXT BOX

16. How long did this work take, from planning to completion?
Please enter the number of months

17. Was the time taken: [dropdown]

Longer than expected
As expected
Shorter than expected

A conditional question is set up for the answer selected to ask ‘Why do you think that was?’

Page 8

What next?

18. Was the future implementation or uptake of the core outcome set considered by your group at any stage? [dropdown]

Yes
No

A conditional question is set up for the answer selected to ask ‘If not, could you please explain why this wasn't a consideration by your group?’ Or ‘If yes, what plans did/do you have to promote the uptake of the core outcome set?’

19. Do you have any plans to update or review your core outcome set? [dropdown]

Yes
No

A conditional question is set up for the answer selected to ask ‘If not, could you please explain why your group doesn't have any plans to update or review the core outcome set?’ Or ‘If yes, what plans did/do you have?’

END OF SURVEY

Appendix 6: COS developers' responses to the survey question 'What do you think were the main strengths of your study?'

Respondent	Strength listed	Category
S29	consensus of experts	Consensus
S35	Consensus was strong	Consensus
S25	development and application of small group consensus process, at that time (1992) innovative	Consensus
S52	consensus process	Consensus
S30	All participants agreed on the terms by which consensus would be met	Consensus
S27	overall consensus	Consensus
S26	structured consensus finding	Consensus
S4	that we achieved some degree of consensus	Consensus
S45	High level of agreement with proposals on outcomes that should be considered	Consensus
S5	international consensus	Consensus
S51	international consensus	Consensus
S12	international consensus	Consensus
S25	in the end, world-wide consensus reached	Consensus
S42	study design (systematic review, consensus panel)	Consensus
S65	first consensus in the field	Consensus
S28	unbiased, statistic based definition of cut offs	Consensus definition/parameters
S50	clear definitions of parameters	Consensus definition/parameters
S35	Wide dissemination and update	Dissemination
S75	Supported and published by the International Society of [disease area]	Endorsement
S47	endorsement by major national organization	Endorsement
S16	Potential for regulatory endorsement	Endorsement
S60	Generally accepted	Endorsement
S18	International representation and buy-in	Endorsement
S62	Provides evidence-based rationale for the necessity of the variables in our recommended core	Evidence based
S20	based on review of literature	Evidence based
S44	Conceptual framework from the initial qualitative research	Evidence based
S39	rigorous, systematic, hypothesis driven, evidence based study	Evidence based
S61	Based on robust reviews of the evidence	Evidence based
S54	own experience was compared to published data	Evidence based
S78	Developed based on own experience and literature	Evidence based

S78	Developed based on own experience and literature	Experience based
S67	based on clinical trial experience	Experience based
S11	The experience I had had with coordinating clinical trials	Experience based
S14	applying validated measures to the constructs	How to measure
S72	Agreed on uniform criteria to assess response to therapy	How to measure
S75	Only using objective outcome measures	How to measure
S79	Analysis of baseline data to assess confluence of measures and outlier/poor measures	How to measure
S79	Looking at additional factors: patient reported outcomes, clinically meaningful change vs. statistically significant change	How to measure
S79	Documented diagnostic strengths and weaknesses, representative sample selection needs, which PRO measures were appropriate for endpoints	How to measure
S65	focus on feasibility	How to measure
S60	Robust measures	How to measure
S2	uniformity in endpoint definition	How to measure
S5	recommendations for directions to pursue for more sensitive and specific outcomes measures	How to measure
S70	We highlighted that many endpoints are poorly defined and have different definitions	How to measure
S81	recognition that consensus definitions of [disease name] are the beginning not the end of the (re)definition pathway	How to measure (definitions)
S77	Clear definitions	How to measure (definitions)
S53	Had a significant impact on future outcomes research in [disease name]	Impact
S61	It has had an enduring impact	Impact
S36	international experts	International
S36	international patients	International
S64	International	International
S61	International - very large group of stakeholders and considered international variations in language and definition	International
S5	international consensus	International
S51	international consensus	International
S3	International expertise	International
S24	International in representation	International
S18	International representation and buy-in	International
S7	International	International
S73	Pan European International Perspective	International
S12	international consensus	International

S38	Reflects an international view of researchers in the area.	International
S23	Breadth of participants (North Am, Europe, PTs, MDs)	International
S62	Representative body of international experts	International
S75	Agreement by international experts	International
S1	based on many discussions with international	International
S66	International perspective (WHO)	International
S48	International agreement on core outcome measures	International
S25	in the end, world-wide consensus reached	International
S8	international perspective	International
S65	international	International
S28	International participation	International
S30	Only surgeons and therapists of established international repute were recruited for the study	International
S6	Representation by international experts	International
S73	Comprehensive review of existing evidence	Lit review
S76	review of the literature	Lit review
S56	two recent systematic reviews of outcome measures	Lit review
S70	We partly knew the literature	Lit review
S73	Comprehensive review of existing outcome measure classifications	Lit review
S42	study design (systematic review, consensus panel)	Lit review
S9	comprehensive review	Lit review
S54	Leading experts performed a search of the literature	Lit review
S13	Literature data	Lit review
S52	literature review	Lit review
S55	Wide literature review	Lit review
S69	Complete review of the whole field (Clinical trials in male sexual dysfunction-MSD)	Lit review
S30	Method circumvented negative group dynamics	Method
S42	study design (systematic review, consensus panel)	Method
S21	used formal nominal group technique	Method
S25	development and application of small group consensus process, at that time (1992) innovative	Novel
S48	First attempt in the field	Novel
S39	first study of its kind: multi-site, multiple modalities	Novel
S53	First comprehensive summary of the state of the art	Novel
S5	first effort to systematically review and critique [disease name] trials outcomes	Novel

S65	first consensus in the field	Novel
S67	new	Novel
S70	no-one had addressed the topic before in [disease name]	Novel
S33	novel	Novel
S71	Novelty in the particular field	Novel
S39	first Class 1 level study	Novel
S14	ICF as framework	Outcomes framework
S10	attempt to map to other disease Common Data Elements	Outcomes framework
S44	Patient research partners were involved at every stage of the study	Patient research partners
S62	Strong methodological process	Process
S1	based on statistical methods the most valid variables were chosen	Process
S1	variables were based on large database with reliable clinical practice data;	Process
S35	Open discussion and debate among all present	Process
S63	lengthy consultation	Process
S63	number of revisions (22)	Process
S25	mix of quantitative and qualitative input	Process
S25	exercises at conference to develop responder index	Process
S61	Modified nominal group technique was rich	Process
S61	It was completed efficiently, and cheaply	Process
S51	listed pros and cons of each approach	Process
S18	Kept list of domains and tools short	Process
S41	Good coordination	Process
S22	Delphi panel process	Process
S78	Based on CONSORT recommendations	Process
S10	significant amount of discussion was involved, high degree of active participation	Process
S50	full statistical model	Process
S26	scientific approach	Process
S9	assessment of many domains	Process
S47	balanced and fair	Process
S24	Addressed practical questions	Process
S41	Well prepared procedures for outcome assessment	Process
S6	Incorporated most current technology	Process
S21	used longitudinal actual data on potential outcome measures	Process
S47	comprehensive	Process - Comprehensive
S74	objective	Process - objective
S7	Robust process	Process - Robust
S62	Addressing an important health issue	Relevance to practice

S30	It did establish what the optimal method for [disease area] correction is	Relevance to practice
S74	relevance of outcomes to clinical practice	Relevance to practice
S6	More relevant to clinical practice, trials, regulatory agencies	Relevance to practice
S27	Reflection of the clinical relevance	Relevance to practice
S6	More relevant to clinical practice, trials, regulatory agencies	Relevance to regulators
S6	More relevant to clinical practice, trials, regulatory agencies	Relevance to research
S15	Relevance to important clinical research studies	Relevance to research
S47	fills critical research gap	Relevance to research
S47	facilitates research excellence	Relevance to research
S50	application to various studies and contexts	Relevance to research
S45	Recognition that not all trials will use same outcome measures as depends on objective of drug being tested	Relevance to research
S61	It was completed efficiently, and cheaply	Resources
S36	high response rate in all three rounds of our e Delphi	Response rate
S28	large number of participants	Sample size
S41	Large size, sufficient statistical power	Sample size
S19	large writing group	Sample size
S10	relatively large group of experts from different training backgrounds	Sample size
S23	Sample Size	Sample size
S61	International - very large group of stakeholders and considered international variations in language and definition	Sample size
S36	lots of patients (compared to previous core outcomes projects)	Sample size
S14	tailored for specific target group	Scope
S45	Need for research on most appropriate outcome measures to be used in different situations	Scope
S77	An internationally known scoring system	Scoring system
S64	All key academic players involved	Stakeholder involvement - academic
S23	Breadth of participants (North Am, Europe, PTs, MDs)	Stakeholder involvement - broad
S66	Broad group of professionals/researchers	Stakeholder involvement - broad
S49	Broad range of experts and professionals	Stakeholder involvement - broad
S15	Broad representation of stakeholders	Stakeholder involvement - broad
S44	A national survey captured a broader patient perspective	Stakeholder involvement - broad

S35	Diverse group of clinical investigators participated	Stakeholder involvement - broad
S34	wide participation - academia, industry, NCI, FDA	Stakeholder involvement - broad
S53	Included a broad perspective	Stakeholder involvement - broad
S41	Multidisciplinary team of scientists working closely together	Stakeholder involvement - broad
S55	Multidisciplinary involvement of authors	Stakeholder involvement - broad
S22	Multidisciplinary panel	Stakeholder involvement - broad
S7	Multiprofessional	Stakeholder involvement - broad
S26	different perspectives included	Stakeholder involvement - broad
S10	relatively large group of experts from different training backgrounds	Stakeholder involvement - expertise
S17	the multidisciplinary team that reviewed the data	Stakeholder involvement - expertise
S79	Strong committee members/author who brought different skills and focus	Stakeholder involvement - expertise
S20	expertise of panel members	Stakeholder involvement - expertise
S66	Experts with large experience in the particular field	Stakeholder involvement - expertise
S9	update with experienced reviewers	Stakeholder involvement - expertise
S34	experienced group	Stakeholder involvement - expertise
S8	Inclusion of different specialists	Stakeholder involvement - expertise
S8	inclusion of experts in the field	Stakeholder involvement - expertise
S51	subject matter world experts participating	Stakeholder involvement - expertise
S3	Multidisciplinary	Stakeholder involvement - expertise
S24	Key experts involved	Stakeholder involvement - expertise
S37	Involvement of a deeply knowledgeable group combining major academic [disease area] researchers and expert clinical research methodologists.	Stakeholder involvement - expertise
S19	experienced writing group	Stakeholder involvement - expertise
S13	Panellists skills and experience	Stakeholder involvement - expertise
S56	group of experts with research and clinical experience	Stakeholder involvement - expertise

S56	mixed qualitative and quantitative researchers	Stakeholder involvement - expertise
S21	panel of experts all had experience both in treating [disease name] but also in doing interventional studies in [disease name]	Stakeholder involvement - expertise
S14	stakeholder involvement	Stakeholder involvement - general
S40	A strong group of authors	Stakeholder involvement - general
S43	Collaboration with bioinformatics groups.	Stakeholder involvement - general
S25	inclusion of industry scientists as full participants from the start	Stakeholder involvement - industry
S37	The study reflected input from leading industry representatives.	Stakeholder involvement - industry
S70	We had someone from the pharmaceutical industry who also have an interest in endpoints	Stakeholder involvement - industry
S66	Broad group of professionals/researchers	Stakeholder involvement - multiple stakeholder groups
S49	Broad range of experts and professionals	Stakeholder involvement - multiple stakeholder groups
S34	wide participation - academia, industry, NCI, FDA	Stakeholder involvement - multiple stakeholder groups
S37	Involvement of a deeply knowledgeable group combining major academic [disease name] researchers and expert clinical research methodologists.	Stakeholder involvement - multiple stakeholder groups
S56	group of experts with research and clinical experience	Stakeholder involvement - multiple stakeholder groups
S21	panel of experts all had experience both in treating [disease name] but also in doing interventional studies in [disease name]	Stakeholder involvement - multiple stakeholder groups
S36	different stakeholders groups	Stakeholder involvement - multiple stakeholder groups
S24	Involvement of multiple stakeholders -regulatory, industry, academics, patient societies	Stakeholder involvement - multiple stakeholder groups
S65	different perspectives included	Stakeholder involvement - multiple stakeholder groups
S44	The methods chosen allowed a patient-centred approach throughout	Stakeholder involvement - patients
S62	Representative body of international experts	Stakeholder involvement - representative

S24	Key experts involved	Stakeholder involvement - KOL
S28	key opinion leaders involved	Stakeholder involvement - KOL
S44	Prioritisation was systematically undertaken with patients	Stakeholder involvement - patients
S36	lots of patients (compared to previous core outcomes projects)	Stakeholder involvement - patients
S48	patients involved	Stakeholder involvement - patients
S63	experts involved	Stakeholder involvement - expertise
S68	Everyone in the field participated	Stakeholder involvement - general
S37	The study reflected input from a wider conference called to address the specific issues covered.	Stakeholder involvement - general
S47	facilitates research excellence	Standardisation
S31	Attempt to define a standard of measurement	Standardisation
S9	proposal of core assessment	Standardisation
S78	Proposed set of clinical outcome indicators	Standardisation
S10	attempt to come up with a battery that could be done in a feasible time frame	Standardisation
S50	standardization	Standardisation
S27	Standardization	Standardisation
S72	Agreed on the based metrics of success and failure that need to be reported	Standardisation
S14	2 existing high profile core sets as basis for development	Starting point
S43	Refinement and extension of prior work.	Starting point
S24	Timely report	Timely
S42	timeliness	Timely
S48	used in subsequent clinical trials, allowing metanalyses	Uptake
S60	Widely adopted	Uptake
S12	prospective validation	Validation
S72	Agreed on how specific patient populations are defined	Wider trial design
S79	Suggested change to duration of trial - initial response followed by maintenance of effect	Wider trial design
S45	To raise awareness about all the issues that should be considered when planning trials in [disease area]	Wider trial design
S46	We established a definition of who to include in child and adolescent [disease name] studies	Wider trial design
S39	manualized therapy that was disseminated to other sites	Wider trial design
S39	demonstrated significant effect size of therapy, despite small sample size	Wider trial design

S81	recognition of the empiricist vs rationalist basis for intervention (and trial) strategies	Wider trial design
S81	recognition that [disease name] trials are never targeted to an individual's phase of stage of injury	Wider trial design
S81	recognition that RCTs were not the only useful trial strategy	Wider trial design
S74	measurable in large scale studies	Wider trial design
S50	consideration of the long-term exposure and risk	Wider trial design
S45	High agreement with criteria to consider for eligibility and management of other drugs during trials	Wider trial design
S79	Documented diagnostic strengths and weaknesses, representative sample selection needs, which PRO measures were appropriate for endpoints	Wider trial design

Appendix 7: COS developers' responses to the survey question 'What do you think were the main challenges of your study?'

Respondent	Challenge	Category
S7	Accessing participants;	Accessing participants
S4	achieving that consensus	Achieving consensus
S35	Balancing the size of the group involved with achieving consensus in brief time available	Achieving consensus
S9	consensus inbetween authors	Achieving consensus
S5	developing consensus	Achieving consensus
S25	no experience in consensus processes of this scope and sensitivity;	Achieving consensus
S14	getting consensus	Achieving consensus
S24	lack of agreement between different regulatory agencies;	Achieving consensus
S68	reaching consensus	Achieving consensus
S43	Reaching consensus.	Achieving consensus
S37	The guidelines included an efficacy effect size criterion, for which consensus proved difficult to achieve.	Achieving consensus
S54	the main challenge was to reach a consensus on standards and definitions	Achieving consensus
S42	reaching consensus	Achieving consensus
S65	some participants insisted in the instruments they developed and used	Bias
S63	expert biases	Bias
S1	to deal with the different (hidden)agenda's/motivations of the professionals	Bias
S6	Old methods ingrained	Changing practice
S60	Changing practice is difficult	Changing practice
S65	established routines in some centers	Changing practice
S65	some participants insisted in the instruments they developed and used	Conflict of interest
S1	to deal with the different (hidden)agenda's/motivations of the professionals	COI
S6	Differences in opinions among the main authors	Differences in opinion
S31	Disagreement among experts	Differences in opinion
S72	Some disagreement among experts	Differences in opinion
S24	lack of agreement between different regulatory agencies;	Differences in opinion
S18	Smoothing political differences	Differences in opinion
S51	differing opinions	Differences in opinion
S60	Experts are disinterested in others views	Differences in opinion
S62	Collaboration among international colleagues	Different interests

	with diverse interests	
S67	different interests between industry and clinicians	Different interests
S20	individuals who focus in biological versus behavioural measures of drug use	Different interests
S48	Dissemination	Dissemination
S35	Balancing what is important with what is feasible in a trial - patient, investigator burden	feasible vs important
S62	Making the work relevant- we worked hard to hone the list to a manageable realistic' level	feasible vs important
S9	find all relevant articles	Finding relevant articles
S63	heterogeneity	General
S45	Lack of gold standard outcome;	General
S60	Conclusions are too general - overlook minority views	Generalisability
S74	patient reported outcomes	How to measure
S50	uncertainties of some of the laboratory methods	How to measure
S76	Upon literature review it was clear there were minimal validated studies/outcomes	How to measure
S10	the same kinds of measures do not uniformly apply to injury severity or chronicity or age across the spectrum, so this was addressed in later efforts;	How to measure
S10	Versions of measures change over time (e.g., WAIS, etc.) which may necessitate substitution. Validation/evidence base vs availability/latest;	How to measure
S10	The same kinds of measures were not always used across different populations (e.g., military vs civilian, sports-related versus other, etc.);	How to measure
S10	considerations pertaining to international use (availability of measures in different languages, cultures, settings)	How to measure
S73	heterogeneity in ascertainment of outcomes across studies;	How to measure
S45	Relative Lack of trials with validated instruments to discuss;	How to measure
S45	Lack of experience of some participants with some potential outcome measures ;	How to measure
S70	Difficult to access the protocols of randomised trials to find how endpoints were actually defined and they were not always defined in the papers	How to measure (definitions)
S28	implementation	Implementation
S24	timing of implementation	Implementation
S37	The American Academy of Periodontology felt that our group, and the American Dental Association, were usurping its prerogative.	Implementation
S15	Organizing stakeholders and including all views	Including all views

S5	lack of available data	Lack of data
S71	Lack of data	Lack of data
S13	No full data available	Lack of data
S8	expert opinion with little data	Lack of data
S73	lack of solid long term mortality data to correlate to nonfatal outcomes	Lack of data
S34	insufficient data in many areas	Lack of data
S70	Lack of data from randomised trials	Lack of data
S72	Lack of good data for some metrics;	Lack of data
S75	Lack of published evidence to validate outcomes	Lack of data
S24	Lack of sufficient longitudinal/natural history data for all outcomes discussed;	Lack of data
S64	Limited literature to support recommendations	Lack of data
S51	weak evidence base;	Lack of data
S49	Weakness of primary research	Lack of data
S79	Weaknesses in the literature, especially bias for specific theoretical models;	Lack of data
S23	Language difficulty;	Language
S61	Understanding across borders;	Language
S48	Multiple domains involved in disability	Multiple domains important
S25	no experience in consensus processes of this scope and sensitivity;	No experience/knowledge of COS development
S48	No previous experience in the field	No experience/knowledge of COS development
S40	limited knowledge in the field	No experience/knowledge of COS development
S7	Absence of methodological guidance	No guidance
S67	never done before in this disorder	Novel
S25	first of its kind, so everything was 'new', many things developed 'on the fly';	Novel
S8	expert opinion with little data	Opinion vs data
S25	finding the balance between expert opinion and data	Opinion vs data
S35	Balancing what is important with what is feasible in a trial - patient, investigator burden	Participant burden
S65	heterogeneity of participants	Participants
S45	Role of comorbidities confounding physician and patient related outcomes	Participants
S36	keep participants motivated and chase them up to reply on questionnaires	Participants - motivation
S44	How to label the outcomes important to patients - to use their own language or map onto existing outcomes?	Patient involvement

S36	explain outcomes and why this is important to non clinical participants (patients)	Patient involvement
S70	we did not have patient perspective	Patient involvement
S16	No patients involved	Patient involvement
S39	patients with seizures cannot drive and may have to come from a distance for treatment	Patient involvement
S40	poor understanding of biology of disease	Poor understanding of disease
S44	Some patients found it difficult to prioritise a large number of outcomes	Prioritising outcomes
S22	Narrowing down the potential research outcomes to a small enough list	Prioritising outcomes
S12	data collection in a large number of centres (>100)	Process
S41	Differences between trial sites in an international multicentre study	Process
S36	keep participants motivated and chase them up to reply on questionnaires	Process
S23	Ensuring that professional organizations distributed the survey to their members	Process
S18	Teaching participating investigators about methods;	Process
S26	to structure the process;	Process
S30	valid interpretation of responses ;	Process
S30	limiting responses to concise statements disallowed complex discussion	Process
S42	volume of effort/review	Process
S55	Open ended question so no limit to the studies that could be included - difficult to know what to include and what not to	Process
S15	Organizing stakeholders and including all views	Process
S63	difficulty arranging meetings (annual)	Process - Arranging meetings
S62	Making the work relevant- we worked hard to hone the list to a manageable realistic' level	Process - list
S22	Narrowing down the potential research outcomes to a small enough list	Process - list
S79	initial and final work done at a distance - collaborative meeting sponsored by ICSM led to specific recommendations from our discussions over 4 days;	Process - remotely
S27	Definition of critical questions	Process - Wording
S44	How to label the outcomes important to patients - to use their own language or map onto existing outcomes?	Process - wording
S81	simplifying the outcomes	Process (simplifying outcomes)
S66	Limited space to present the review (in conference proceedings)	Publishing

S59	Getting subjects early in their disease course	Recruitment
S29	quick development of treatment options	Keeping up to date
S11	Covering more recent treatment areas	Keeping up to date
S21	Y2K wiped out our longitudinal database of clinical measures	Resources - IT
S26	funding	Resources - funding
S3	Insufficient resources for systematic review	Resources - general
S66	Limited time/resources to undertake the review	Resources - general
S7	Managing Delphi process through ICT;	Resources - IT
S47	financial limitations for support of experts;	Resources - money
S35	Balancing the size of the group involved with achieving consensus in brief time available	Resources - time
S47	time/availability of experts	Resources - time
S66	Limited time/resources to undertake the review	Resources - time
S43	Time required to complete the work.	Resources - Time
S56	staff time	Resources - time
S26	lengthy process;	Resources - Time
S61	Speed meant we made a few mistakes	Resources - Time
S10	we did not have the time to do an EXHAUSTIVE literature review or coding of evidence base level;	Resources - Time
S23	Response rate;	Response rate
S35	Balancing the size of the group involved with achieving consensus in brief time available	Size of the group
S79	Small number of contributors - each had a lot of work to do	Size of the group
S36	ensure the balance of clinicians and patients in the study is the same	Size of the group(s)
S28	further validation	Validation
S44	Some outcomes e.g. activities of daily living and independence may have a strong degree of overlap - testing required to create a robust core set	Validation
S53	We realized that most outcomes had not been validated at that time.	Validation
S46	Defining the age of onset and the definition of mania	Wider trial design
S45	Variety of types of trials that could be done in this single disease;	Wider trial design
S78	use CONSORT guidelines for outcomes in surgical trials	Wider trial design
S17	we were to develop protocols to study drugs that had not had appropriate studies done in newborns	Wider trial design

Appendix 8: COS developers' responses to the survey question 'What do you think were the main limitations of your study?'

Respondent	Limitation	Category
S42	expert panels always have limitations (bias)	Bias
S63	expert biases	Bias
S5	the heterogeneity of [disease name] and the need for different outcomes at different times and for different phenotypes of the disease	Disease/population
S24	Lack of accurate estimate of pediatric [disease name] population	Disease/population
S77	The complexity of the particular patient group	Disease/population
S54	the proposed definitions have not been backed up by the corresponding scientific societies	Endorsement
S61	We have not canvassed the research community to see what they have thought about the consensus, although it is widely quoted	Endorsement
S10	the initial publication acknowledged a weakness in pediatric measures, so this was addressed in future efforts	How to measure
S21	PROs were in very early stage of development	How to measure
S21	we went for response measure and should have also developed a state measure at the same time	How to measure
S45	Lack of instruments validated in clinical trials to use in future trials ;	How to measure
S45	Most outcome measures validated in observational studies;	How to measure
S48	Low responsiveness due to slow disease evolution	How to measure
S52	not all domains had instruments available	How to measure
S65	limited data about change sensitivity of instruments	How to measure
S10	the project considered the application of these measures in populations of patients, but not necessarily in the selection of control groups for these	How to measure
S15	Limited to endpoint definitions and did not include analysis methods	How to measure
S17	no known norms for blood pressure in the very premature infant, that had not been determined	How to measure
S41	Statistical power to address subgroup analyses;	How to measure
S41	Limited attention to rigorous assessment of patient-centered outcomes	How to measure
S4	feasibility remains an unanswered question	Implementation
S38	The recommendations will not necessarily be followed.	Implementation
S44	Following on from 1) - this means that it is harder to get researchers and clinicians using the patient core set;	Implementation

S67	that despite firm recommendation industry developed own outcome	Implementation
S72	A pre and post analysis has not been done to determine success of our guideline;	Implementation
S75	Lack of ability to ensure that these outcomes will be used in clinical trials	Implementation
S37	The work was sound, but the American Academy of [disease area] felt compelled to in essence redo and overwrite it with its own document.	Implementation
S5	maintaining the difference between needs for trials outcomes (i.e. research) versus practice outcomes (i.e., monitoring disease in daily practice)	Intended use (Trials vs practice)
S21	performance of the core set and calculation of response not easily adapted to routine clinical care	Intended use (Trials vs practice)
S5	lack of data to serve as an initial resource	Lack of data
S9	not confronted with studies	Lack of data
S12	lack of performance in other data set	Lack of data
S12	lack of clinical trials in juvenile dermatomyositis and lupus	Lack of data
S13	Lack of comparative trials among some drugs	Lack of data
S29	missing randomized studies	Lack of data
S34	insufficient data in many areas;	Lack of data
S34	evidence base lacking for some decisions	Lack of data
S41	Limited breadth of data on risk factors;	Lack of data
S45	Heterogeneity of previous trials that influenced discussions;	Lack of data
S3	Lack of systematic review	Lack of data
S6	Not always data driven	Lack of data
S6	Unable to include data in development not validated	Lack of data
S24	Lack of sufficient longitudinal/natural history data for all outcomes discussed;	Lack of data
S43	Only a portion of the knowledge base included.	Lack of data
S64	Limited evidence behind recommendations	Lack of data
S66	Seen in a historic perspective, the underlying literature search was limited;	Lack of data
S70	Lack of data from randomised trials ;	Lack of data
S70	Randomised trials were not mature;	Lack of data
S70	randomised trials were historical and contemporary treatment had improved;	Lack of data
S76	Upon literature review it was clear there were minimal validated studies/outcomes	Lack of data
S11	It was based upon my experience - as indicated it was not a study, but a review	Lack of data (based on experience/opinion)
S27	More consensus-based than evidence-based	Lack of data (based on experience/opinion)

S28	at the end it is expert opinion, not evidence	Lack of data (based on experience/opinion)
S45	Conclusions based more on expert view than trial data;	Lack of data (based on experience/opinion)
S68	Expert consensus approach	Lack of data (based on experience/opinion)
S56	it was not designed as a study, rather a consensus from a group of experts - pragmatic approach	Lack of data (based on experience/opinion)
S30	Some eminent surgeons and therapists could not participate due to language difficulties	Language difficulties
S23	Language difficulties	Language difficulties
S40	limited knowledge in the field	Limited knowledge
S60	Too focused on inflammation;	Narrow focus
S61	We left out head injury when we should have included it and will update the measures;	Narrow focus
S62	The findings are applicable to our population	Narrow focus
S2	no golden standard available	No gold standard
S46	There is no gold standard to define those things	No gold standard
S45	Results were not very specific but raised awareness about points to consider in trial design and use of outcome measures that are validated	Non-specific results
S59	could have enrolled more subjects	Number of participants
S28	too few cardiologists;	Number of participants
S39	small sample size	Number of participants
S65	relative small number of experts;	Number of participants
S36	although lots of patients replied to our project, their number was still lower than clinicians, hence possibly they were underrepresented	Number of participants/Representativeness
S44	Despite efforts to recruit from minority groups, these were not well represented	Number of participants/Representativeness
S50	complexity of the situations to apply the methodology	Other
S53	It was not a study	Other
S60	Over-emphasize drug effects;	Other
S79	Specific agendas by some experts not included in this chapter;	Other
S5	ultimately, the inability to recommend a core outcome set, but rather limited to critiquing those commonly used outcomes at the time (1994/5)	Inability to recommend a COS

S20	rather than a single core set of measures to use across studies, aspects of the study will determine which core measures from among a larger set will be used	Scope (study design)
S31	Clinically unproven	Lack of validation of COS
S33	lack of preceding consensus	Lack of preceding consensus
S41	Differences between centres;	Others
S9	new studies have new challenges	Out of date (keeping current/updated)
S15	Dynamic field where outcomes of interest change	Out of date (keeping current/updated)
S21	newer outcome measures and domains have developed since then	Out of date (keeping current/updated)
S48	Need to update ;	Out of date (keeping current/updated)
S51	outdated at this point;	Out of date (keeping current/updated)
S51	needs an update	Out of date (keeping current/updated)
S73	ever changing standard of care	Out of date (keeping current/updated)
S6	Did not address all of the intended issues	Process
S30	It did not allow for innovation;	Process
S35	Did not provide a forum for external input to the consensus process	Process
S55	Each study was not reviewed in a standardised way,;	Process
S55	Only English language literature	Process
S70	Not sure we captured as many studies as we should in our searches because of different nomenclature	Process
S71	Nominal group methodology	Process
S79	Much of the work was done while authors continued to do their full time work	Process
S23	Response rate;	Response rate
S22	we were not able to consider procedures other than UAE	Scope (intervention)
S7	We need to reduce size of COS	Size of COS
S26	even more perspectives needed, e.g. health authorities	Stakeholders (general)
S74	only professionals involved	stakeholders (general)
S47	some experts who might have facilitated had other commitments	Stakeholders (general)
S72	Did not include representatives from industry	stakeholders (Industry involvement)
S78	view of other international experts also needed;	stakeholders

		(international)
S1	limited patient involvement	stakeholders (Patient involvement)
S8	lack of patient input	stakeholders (Patient involvement)
S16	no patients involved	stakeholders (Patient involvement)
S18	Lack of direct patient input (we are addressing that now)	stakeholders (Patient involvement)
S25	no patient involvement (solved in 2002);	stakeholders (Patient involvement)
S35	Did not include patient advocates;	stakeholders (Patient involvement)
S36	although lots of patients replied to our project, their number was still lower than clinicians, hence possibly they were underrepresented	stakeholders (Patient involvement)
S44	It would have been preferable to include the patient perspective at the time of the original development of the core set rather than later;	stakeholders (Patient involvement)
S44	Following on from 1) - this means that it is harder to get researchers and clinicians using the patient core set;	stakeholders (Patient involvement)
S49	Unable to include forensic mental health service users	stakeholders (Patient involvement)
S56	if designed as a study would include a greater range of stakeholders including patients and families	stakeholders (Patient involvement)
S60	Overlook important benefits to patients	Stakeholders (Patient involvement)
S61	We did not have a multi-layered patient involvement process, we selected measures that had been developed through patient involvement;	stakeholders (Patient involvement)
S66	No patient representative;	stakeholders (Patient involvement)
S70	No patient perspective;	stakeholders (Patient involvement)
S25	limited regulatory involvement	stakeholders (Regulatory involvement)
S9	theoretical proposition	Theoretical - discuss with lack of data?
S66	A theoretical overview and not an empirical study of papers/reports	Theoretical - discuss with lack of data?
S81	conceptual not a trial	Theoretical - discuss with lack of data?
S74	no validation of outcomes data set	Validation

Appendix 9: COS developers' responses to the survey question 'Reflecting on your experiences of developing a core outcome set, are there any areas that you feel would benefit from methodological guidance or research to inform future activity to develop core outcome sets?'

Respondent	Guideline requirements	Category
S64	The ability to deliver international registries	Application of COS
S7	what is consensus	Consensus methods
S12	consensus formation techniques	Consensus methods
S44	Noticed with others developing core sets - lots of confusion about consensus levels for Delphi/ NGT	Consensus methods
S27	Structured method for selection of experts and reaching consensus	Consensus methods
S25	We've now developed OMERACT Filter 2.0 that is highly relevant to all COS developers. This, or something strongly like it, should be adopted by COMET ASAP. COS developers need concrete guidance NOW	Consensus methods
S7	homogenous versus heterogeneous samples;	Consensus methods
S23	Selecting potential outcomes	Consensus methods
S42	including other methodology experts to observe for potential bias;	Consensus methods
S18	Better definitions of what domains and sub-domains are in this context	Outcome terminology
S38	There could be general guidelines about defining types of outcomes.	Outcome terminology
S7	All aspects;	General
S31	Absolutely, any additional guidance possible offers best hope of developing appropriate guidelines	General
S36	systematic approach;	General
S37	The guidelines development effort was led by research methodologists, most specifically myself, and had ample external methodological advice available	General
S54	probably, this is not my main focus of expertise anymore	General
S61	I don't think it is rocket science, but guidance would be good;	General
S61	I think guidance needs to reflect the practicalities of making decisions based on imperfect information;	General
S61	I don't think that guidance should be too constraining, there are many ways to skin a cat;	General
S63	probably	General
S66	This field is to me a moving target with constant need of overview and deliberations;	General

S67	guidelines for optimizing the process	General
S65	for a next step methodological guidance might be useful	General
S10	use of evidence-based guidelines	General
S4	we don't know what a clinically meaningful effect size will be	How to measure
S5	there still (nearly 20 years later) needs to be informative and useful data on such issues as biomarkers (lab-based and imaging)	How to measure
S5	more data still needed on interface between MRI and clinical outcomes	How to measure
S5	more flexibility on the part of regulators on surrogacy and use of MRI as a surrogate outcome	How to measure
S6	More quantitative than qualitative assessments	How to measure
S21	response measure versus status measure choice in design stage	How to measure
S45	Having more clinical trial data to assess how individual outcome measures and composite indices worked would help;	How to measure
S60	Better generic health status measures;	How to measure
S60	Simpler ways to capture cost-effectiveness	How to measure
S70	PROMS	How to measure
S74	integrating clinic and patient perspectives into composite outcomes	How to measure
S76	The outcome measures need to be validated first,;	How to measure
S76	Then you might be able to develop a composite outcome	How to measure
S77	There are several severity scores in use;	How to measure
S77	The outcome death/survival may be same	How to measure
S77	but risk predictions may differ and level of risk;	How to measure
S77	may be consequence of the particular severity score.	How to measure
S48	Advice on how to make regulatory agencies use the core set	Implementation
S66	The relation/balance between (masses of) core outcomes and specific results related to the hypothesis tested sometimes difficult	Implementation
S72	Pre and post analysis	Implementation
S47	ensure endorsement mechanism is in place in advance	Implementation
S8	define different approaches and pros and cons of each	Methods
S48	Advice on choice of method;	Methods
S51	gleaning methods from other diseases	Methods
S62	Focus group methods;	Methods
S78	Yes. Methodology recommendations, like those available for guideline development	Methods
S36	which method to use (Delphi, focus groups etc);	Methods

S34	more data	More data
S36	what are the criteria for consideration of a core outcomes project as valid	Quality assessment
S35	Need to demonstrate the validity and utility of the recommendations;	Quality assessment
S43	Constant review, refinement, and updating.	Review and feedback
S3	Systematic review	SR
	include other librarian to shoot literature search separately and look for degree of agreement	SR
S55	Guidance on developing a clear rationale for study inclusion	SR
S23	Inviting participants to act as stakeholders in the development of the recommendations.	Stakeholder involvement
S27	Structured method for selection of experts and reaching consensus	Stakeholder involvement
S30	The choice of participants for a Delphi study should be made by a group of independent observers ;	Stakeholder involvement
	Independent observers that should be tasked to choose Delphi participants should be guided by patients	Stakeholder involvement
S32	Input from multiple perspectives	Stakeholder involvement
S35	Involvement of external groups to increase acceptance	Stakeholder involvement
S36	how to identify participants;	Stakeholder involvement
S38	There could be general guidelines about defining participant populations and recording of sociodemographic factors.	Stakeholder involvement
S39	having a patient group that discusses relevant outcomes	Stakeholder involvement
S44	More discussion about including minority patient participants in the process;	Stakeholder involvement
S45	Need international panel from worldwide research community with relevant experience of undertaking and analysing trial data	Stakeholder involvement
S48	Advice on choice of participants and stakeholders;	Stakeholder involvement
S51	increase input from patients and patient advocacy organizations	Stakeholder involvement
S62	How to better incorporate patients at an appropriate level and time	Stakeholder involvement
S73	need to incorporate patient perspective	Stakeholder involvement
S79	Include individuals from interested agencies FDA/EMA, NIH	Stakeholder involvement
S44	Guidance on wording of outcomes and domains when including patients in the process;	Wording

Appendix 10: Relevant qualitative training

Introduction to Qualitative Interviewing course; completed 2012 (University of Oxford)

Analysing Qualitative Interviews course; completed 11-12 April 2013 (University of Oxford)

MEDR628: Epistemological and Methodological Approaches to Qualitative Research on Medicine, Health & Society 12 week lecture series; completed 2013 (University of Liverpool)

MEDR629: Qualitative Research Methods – Methods workshops; completed July 2013 (University of Liverpool)

Introduction to NVivo for qualitative data analysis; completed March 2014 (University of Liverpool)

Appendix 11: Version 1.0 of the interview topic guide

Topic Guide for interviewing researchers

This topic guide outlines the questions and prompts that may be used during the interviews. It may be adapted as the study proceeds according to participant's needs and responses, as well as the emerging analysis.

Introduction

Explanation of the context of the study and terminology being used (emphasise that the focus is on 'what' part of core outcome set development not 'how')

Explain that all q's in relation to study/project [title]

Main discussion

Background

1. Tell me about your role in the development of this core outcome set
    - How did you become interested in developing a core outcome set in [clinical area]
    - Had you worked on a core outcome set previously?

2. Tell me about your research [title]
    - What were the main objective/aims?
    - Were there any secondary aims?
        - If so, what were they

3. How did this study come about?
    - What was the motivation
    - What was your motivation for getting involved
    - What was the current situation/environment
        - What was already known
    - Did anything else influence your initial thinking

4. What role did each individual have in the planning of this work
    - Your role in the planning of this work
    - Who else was involved, and what were their roles
    - Were any patient or public research partners involved in the planning of the work?

- If so, tell me about it

Choice of methods

5. Can you talk me through the planning of this project
    - Was there a protocol?
    - Did you consider the scope of the core outcome set (e.g. all interventions, specific intervention etc.)?
        - Was this at the planning stage or come later?
        - How did you make the final decisions on what scope to cover?

6. Was the study funded?

    - Who funded the project?
    - Tell me about the funding
    - How easy/difficult was it to obtain funding for this work?
    - Did any funders turn the proposal down?
    - Did you receive any advice about how to get funding? What? Who from? Was this useful?
    - How did you decide where to apply?
    - Did you use other resources to get the project done?
        - If so, what were they?

7. Can you tell me about the methods you used?
    - How did you decide on the methods to use – why this/these method?
        - Any pilot work?
        - Did the methods differ by stakeholder group?
    - What influenced this approach?
        - Previous work
        - Clinical area
        - Experience of methods
    - Who influenced this
    - Who made the final decision on the methods?
    - Did you consider using any other methods?
        - If not, why?
        - If so, which ones? Why did you decide not to use them?

8. Do you think that the available resources influenced the methods you used?
    - If so, how?
    - If not, why?

9. Did the methods change throughout the process?
    - If so, how?
    - Why do you think that was/wasn't?

10. Did you experience any difficulties in maintaining interest in the process?
    - E.g. if it was lengthy, required input on more than one occasion
    - Consider patient involvement

Choice of stakeholders to include

11. Who was included in the development of this core outcome set?

- *Be clear they understand as participants not as members of research team*

12. How did you decide who to include in this work?
    - E.g. important, resources etc.
    - Was this easy/ difficult?

13. How did you decide how many and the proportions of different types of people to include?
    - What was the response rate to the research team's invitation to participate? (Different for different groups)?

14. Did you end up including the numbers and proportions you intended?

15. How did you explain the study to different groups?
    - Did you have to explain things differently to patients?
    - Were there misunderstandings and how were these dealt with?

16. Did you think about including any groups that you didn't end up including?
    - Key stakeholder groups e.g. patients/PPI, clinical experts, industry representatives, authorities (e.g. governmental, regulatory).
    - Why were they not included?
        - Did you try to include them but they declined, or did you decide not to include them?

17. How did you decide to integrate opinions from different stakeholders?
    - If used a single heterogeneous panel, did you examine differences of opinions between groups?
        - Was this considered?

    - Had you included strategies to cope with disagreements when planning the project?
        - If not, why not?
        - If yes, what strategies did you include?

    - Were there any disagreements?
        - If so, how was this handled?
        - If you got differences what did you do?

- Did these considerations influence the methods you used then?

Process and methods of analysis

18. How did you decide what information (if any) to give to participants before the process?

19. How did you decide on how outcomes would be scored or rated during the consensus exercise or at each stage in the process?

20. Did you adopt a particular definition for consensus?

    - If not, was this considered?
    - If yes, when was the definition decided (for example at the protocol stage, during the rating/scoring or analysis stage?)
        - And what was the rationale behind that definition?

21. How did you decide the procedure for determining how outcomes would be included or excluded from consideration at each stage of the consensus process?

22. How did the researchers ask the question(s) about outcomes?

    - What was the specific wording?
    - Did you do this differently for different stakeholder groups?

Results

23. What did you think about the content of the core outcome set at the end of the study?
    - Were the final outcomes in the COS anticipated by you?
        - Why/why not?
    - Were the final outcomes in the COS anticipated by others involved in initiating the project?
        - Why/why not?
    - Can you tell me about one outcome that was included and you were confident it would be at the outset?
    - Can you tell me about one outcome that was included and you were confident that it would not be at the outset?
    - Can you tell me about one outcome that was not included but which you were confident would be included at the outset?
        - Why do you think these three examples went the way they did?

24. What influenced the final choice of outcomes to include in the final core outcome set?
    - Did you consider having short term and long term and Patient Reported Outcomes all within a COS?

- Do you think it is possible to include these in the same core outcome set?

25. Were there differences in the outcomes identified and thought to be important by different groups?
    - What outcomes did patients identify as important that clinicians didn't?
    - What was the same?
    - What did clinicians say was important that patients didn't?

26. Did the development of this core outcome set change anything?
    - Influence the research community in this area
        - If so, how?
        - If not, why?
        - If definitive yes or no given, what's your evidence source for this?

27. Did the results influence your thinking?
    - If so, how?
    - If not, why not?

Publication

We want to find out about your experience of getting this work published so that we can help others with getting this type of study published.

28. Can you tell me about the publication process for this project
    - Did you use any reporting guidance?
    - How was the decision made regarding choice of journal?
    - Were any problems encountered?
    - Was the paper accepted by first journal?
        - If not, which journal(s)?
        - If not, what reasons were given for this?
    - Did the reviewers/editors have any influence on the final format/content of the paper?
        - If so, how?

General

29. What (if any) have you experienced to be the challenges of this work?

30. What (if any) have you experienced to be the benefits of this work?
    - What do you think they will be?

31. Would you do anything differently?
    - (Additional stakeholders, different choice of methods...)

32. Were there any areas particularly that you felt would benefit from methodological research?
    - Do you think you would have benefited from guidance in developing a COS?
        - If so, what?
        - If not, why?

33. Did you go as far as to consider the implementation of your core outcome set?
    - Can you tell me about it
    - What did you do?
    - What was successful?
    - Have you experienced/anticipate potential barriers to implementation?

34. Do you think there were any limitations to your work?
    - (E.g. stakeholders included, geographical coverage, lack of anonymity etc.)

35. What do you think are the implications for future research?
    - Do you have plans to review the COS?
        - If so, how?
            - Funding
            - Who – individuals, organisations?
            - What would make you decide to start the updating?
        - If not, why not?

36. What impact do you think this work will have?
    - Who are the key groups that should benefit from this work?
    - What do you consider its relevance to these key groups?
        - E.g. patients and the public, healthcare providers, researchers, pharmaceutical companies, device manufactures, policy makers
    - Will you assess this?
        - If so, how?
        - If not, why not?

37. What is it intended the core outcome set is used for?
    a. Do you consider the results of this study to be applicable beyond use in clinical trials? E.g. systematic reviews, clinical practice, clinical guidelines

        - If so, what?
        - If not, why not?

38. Do you think there were any conflicts of interest within the study team?
    - (e.g. researchers who have developed an outcome measurement instrument applicable to the scope of the COS)
    - Or industry wanting shorter term outcomes to be included rather than longer term outcomes (because cheaper?)

39. Do you plan to work on another core outcome set?
    - If you did work on another core outcome set (whether or not you plan to), can you tell me about anything that you would definitely want to do in the same way?
    - And about anything that you would definitely want to do in a different way?

Other issues

40. Are you still in touch with the patients or members of the public included in the study?

41. Would you like to receive a copy of the publication?

That's the end of my questions.

42. Is there anything else that is important to you that we haven't talked about today?

43. Is there anything else you'd like to say?

……

Close interview

Appendix 12: Invite email

Date

Dear [lead author]

I am a PhD student working with Professor Paula Williamson (Department of Biostatistics) and Professor Bridget Young (Department of Psychological Sciences) at the University of Liverpool, UK. I am based within the Medical Research Council Network of Hubs for Trials Methodology Research, and I'm also the Project Coordinator for the COMET (Core Outcome Measures in Effectiveness Trials) Initiative. The COMET Initiative brings together researchers interested in the development and application of agreed standardised sets of outcomes, known as a 'core outcome set.' More information about the COMET Initiative can be found from http://www.comet-initiative.org/.

I'm conducting a study that aims to assist the research community by informing the development of guidance standards for core outcome set development. We have recently completed a systematic review to identify studies that have been conducted to determine the outcomes or domains to measure in clinical trials. The major focus of my study is to improve our methodological understanding by exploring researchers' experiences and views of core outcome set development and implementation.

We read with interest your study [title] published in [journal title] which was included in our systematic review. You have been asked to take part in this study because you are or have been involved in research to determine the outcomes or domains to measure in clinical trials, and your experiences are therefore very important to us. We would like to learn from your experiences and incorporate what we learn into guidance to assist future researchers.

I would be grateful if you would be willing to be interviewed by telephone about your experiences. If you would like more information about the project one of my supervisors or myself would be happy to talk by phone or email. If you are happy to participate, any information you provide would remain anonymous and would be combined with information provided by other researchers who also take part.

Please can you let me know if you would be willing to be interviewed? I will then contact you again, to arrange a convenient time to carry out the interview.

I am most appreciative of your help and I look forward to hearing from you soon.

Yours sincerely

Elizabeth Gargon

****

Elizabeth Gargon

COMET Project Coordinator

Department of Biostatistics, University of Liverpool, 1st floor Duncan Building, Daulby Street,

Liverpool, L69 3GA

e-mail: e.gargon@liv.ac.uk

Tel: +44(0) 151 706 5955

Appendix 13: Participant information sheet

Exploring researchers' experiences and views of core outcome set development and implementation

You are being invited to take part in a research study

You are being invited to take part in a research study. Before you decide whether to take part, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully, and ask if you would like more information or if there is anything unclear or you don't fully understand. Feel free to discuss this with colleagues or others if you wish.

Who is doing the research?

Elizabeth Gargon, PhD student in the Department of Biostatistics at the University of Liverpool, will be carrying out the interviews.

I am a PhD student funded by the Medical Research Council Network of Hubs for Trials Methodology Research. My supervisors are Professor Paula Williamson, Department of Biostatistics at the University of Liverpool, and Professor Bridget Young, Department of Psychological Sciences also at the University of Liverpool.

What is the purpose of the study?

This study will contribute to a larger research project that is aiming to develop methodological guidance for core outcome set development. In order to formulate guidance in this area we need to try to understand what factors have informed the ways in which researchers have developed core outcome sets. We would like to learn from your experiences and incorporate what we learn into guidance to assist future researchers.

Why have I been chosen to take part?

You have been asked to take part in this study because you are or have been involved in research to determine what outcomes or domains to measure in clinical trials. Your experiences are therefore very important to us.

Do I have to take part?

No. Your participation is entirely voluntary. If you decide to take part, you will be asked to sign a consent form. If you decide later on that you wish to withdraw from the study then you can leave up until the point of interview transcript anonymisation, and you do not need to give a reason.

What will happen if I take part?

The interview will involve you speaking to the researcher about your experiences of conducting research to determine what outcomes or domains to measure in clinical trials. The interview will last approximately 45 minutes to an hour. With your permission, the interview will be audio-recorded. You can stop the interview at any time, and you do not have to answer a particular question if you don't want to.

Where will the interview take place?

The interview will be conducted by telephone, at a time and date convenient to you.

Are there any risks in taking part?

We do not expect there to be any risks or discomfort to be associated in this research study. However, if you feel uncomfortable then you can stop the interview at any given time and without providing a reason.

Are there any benefits in taking part?

You will be helping with a new area of research. Your experiences will help us to learn about what methods are most helpful in determining which outcomes or domains to measure in clinical trials. We hope this study will help to improve this process in the future.

What if there is a problem?

If you have a concern about any aspect of this study, please feel free to let us know by contacting the lead researcher, Elizabeth Gargon, who will try to help and answer your

questions. If you remain unhappy or have a complaint which you feel we cannot deal with, then you can contact the Research Governance Officer at the University of Liverpool ethics@liv.ac.uk, 0151 794 8290. When contacting the Research Governance Officer, please provide details of the name or description of the study (so that it can be identified), the researcher(s) involved, and the details of the complaint you wish to make.

Will my participation be kept confidential?

All of the information that you give us will be kept strictly confidential. Only Elizabeth Gargon will have access to any information about you. My supervisor, Bridget Young, may listen to a recording of the interview to advise on my interview technique. The audio-recordings of the interviews will be marked with a number only. These audio recordings will be transcribed but identifying details such as place and person names will be removed from the transcripts. Only details necessary for interpretation (e.g. clinical area) will be kept in. We may use brief quotes from your interview in the write-up of the study. We will ensure that identifiable details (e.g. person and place names) have been removed, although it will be necessary to indicate the general clinical areas in which core outcome set work has taken place. All the information that you provide during the study will be stored in locked filling cabinets and/or password protected computers.

At the end of the study, audio recordings will be destroyed. All other research data (consent forms, anonymised interview transcripts, field notes, and contact details) will be kept in locked filing cabinets and/or password protected university computers for ten years.

Will my taking part be covered by an insurance scheme?

Participants taking part in a University of Liverpool ethically approved study will have cover.

What will happen to the results of the study?

The results will be written up as part of Elizabeth Gargon's postgraduate research thesis and submitted for examination. The findings will also be published in academic journals and presented at conferences. If you wish, you will be provided with a summary of the findings at the end of the study and a copy of the final research report.

What will happen if I want to stop taking part?

If you decide that you no longer want to take part, then you can stop up until the point of interview transcript anonymisation and do not need to give a reason for doing so. Results up to the period of withdrawal may still be used. Otherwise you can ask for your data to be removed from the study and destroyed.

How can I find out more?

Get in touch with the researcher, Elizabeth Gargon:

Elizabeth Gargon

Department of Biostatistics, University of Liverpool, 1st floor Duncan Building, Daulby Street, Liverpool, L69 3GA

e-mail: e.gargon@liv.ac.uk

Tel: +44(0) 151 706 5955

The contact details of my supervisor are:

Paula Williamson

Department of Biostatistics, University of Liverpool, 1st floor Duncan Building, Daulby Street, Liverpool, L69 3GA

e-mail: prw@liv.ac.uk

Tel: +44(0) 151 706 4958

Thank you for reading this information sheet.

This is your copy to keep; please retain for your records.

Appendix 14: Consent form

Exploring researchers' experiences and views of core outcome set development and implementation

Researcher: Elizabeth Gargon

	Please initial box
1. I confirm that I have read and have understood the information sheet dated [DATE] for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.	☐
2. I understand that my participation is voluntary and that I am free to withdraw up until the point of interview transcript anonymisation without giving any reason, without my rights being affected.	☐
3. I understand that, under the Data Protection Act, I can at any time ask for access to the information I provide and I can also request the destruction of that information if I wish.	☐
4. I understand that quotes from what I say during the study may be used where necessary, on the condition that my identity remains anonymous.	☐
5. I agree to the interview being audio-recorded.	☐
6. I agree to take part in the above study.	☐

_____ _____ _____

Participant Name Date Signature

_____ _____ _____

Researcher Date Signature

Please complete this form and return it to e.gargon@liv.ac.uk. The researcher will then sign the form, and a copy of the completed consent form will be returned to you.

The contact details of lead Researcher are:

Elizabeth Gargon

Department of Biostatistics, University of Liverpool, 1st floor Duncan Building, Daulby Street, Liverpool, L69 3GA

e-mail: e.gargon@liv.ac.uk

Tel: +44(0) 151 706 5955

The contact details of my supervisor are:

Paula Williamson

Department of Biostatistics, University of Liverpool, 1st floor Duncan Building, Daulby Street, Liverpool, L69 3GA

e-mail: prw@liv.ac.uk

Tel: +44(0) 151 706 4958

Appendix 15: Summary characteristics of interviewed COS

	Published (n=18) Number of COS	**Ongoing (n=14)** Number of COS	**TOTAL (N=32)** Number of COS
Methods used (not mutually exclusive)			
Systematic/literature review	7	8	**15**
Delphi	7	10	**17**
Semi structured group discussion*	10	6	**16**
Consensus development conference	2	0	**2**
Nominal group technique	1	0	**1**
Focus groups	0	3	**3**
Survey	2	2	**4**
Unstructured group discussion**	1	0	**1**
Interviews	0	5	**5**
Methods not described	1	2^	**3**
People involved			
Clinical expert (no PPI)	9	2	**11**
PPI	8	12	**20**
Not described	1	0	**1**
Study aims			
Outcomes only	11	12	**23**
Wider trial design issues	7	2^^	**9**
How to measure			
What to measure only	8	12	**20**
What to measure + discussion/consideration of how but no recommendation	4	n/a	**4**
What + how (done together)	5	1	**6**
What + how (done in two stages)	1	1	**2**
Population characteristics – Age			
All	1	1	**2**
Children	4	3	**7**
Adults	3	10	**13**
Not specified	10	0	**10**
Intervention characteristics			
All	3	8	**11**
Drug	4	0	**4**
surgery	0	5	**5**
Specific (other)	3	1	**4**
Not specified	8	0	**8**
Number of COS/disease category			
Clinical area where more than 5	14	11	**25**
Clinical area where less than 5	4	3	**7**
Funding			
Commercial	3	0	**3**
Non-commercial	7	13	**20**
Commercial and non-commercial	3	0	**3**
No funding	1	1	**2**
Not reported	4	0	**4**
Plans to review/update			
Yes	1	n/a	**1**
No	17	n/a	**17**
Year of earliest publication			
Pre 2010	4	n/a	**4**
2010-2013	14	n/a	**14**
Ongoing	n/a	14	**14**

**Descriptions included workshop, meeting, and roundtable*
*** Descriptions included task force, work group, working group/party, committee, board, and panel*
^ Consensus methods, in addition to systematic review, were not known prior to interview
^^ Part of a wider PhD

Appendix 16: Copy of publications arising from the work in this thesis

References for the relevant articles are provided and copies included.

Chapter 2

The work contained in chapter 2 has been published in BMC Medical Research Methodology:

Gargon, E., P. R. Williamson, et al. (2015). "Collating the knowledge base for core outcome set development: developing and appraising the search strategy for a systematic review." BMC Med Res Methodology 15(1): 26. Highly accessed.

Chapter 3

The work contained in chapter 3 has been published in PLoS One:

Gargon, E., B. Gurung, et al. (2014). "Choosing important health outcomes for comparative effectiveness research: a systematic review." PLoS ONE **9**(6): e99111.

This text box is where the unabridged thesis included the following third party material:

Gargon, E., P. R. Williamson, et al. (2015). "Collating the knowledge base for core outcome set development: developing and appraising the search strategy for a systematic review." BMC Med Res Methodology 15(1): 26

DOI: 10.1186/s12874-015-0019-9

This text box is where the unabridged thesis included the following third party material:

Gargon, E., B. Gurung, et al. (2014). "Choosing important health outcomes for comparative effectiveness research: a systematic review." PLoS ONE **9**(6): e99111.

DOI: http://dx.doi.org/10.1371/journal.pone.0099111