Improving the Relevance and Consistency of Outcomes Reported in Clinical Trials

Abstract

One of the defining features of patient-centered outcomes research (PCOR) is the emphasis on reporting outcomes that are meaningful to patients. Accelerating progress toward this objective could be achieved through increased development and uptake of core outcome sets (COS), which are intended to represent a standardized minimum set of outcomes that should be measured and reported in all clinical trials in a specific condition. The level of activity around COS has increased significantly over recent years, however there are many important clinical conditions for which high quality COS have not been developed. We believe that meaningful progress toward the goals behind the significant investments in PCOR will depend on sustained attention to the challenges of COS development and uptake.

Introduction

In August 2015, the Patient-Centered Outcomes Research Institute (PCORI) announced that they had surpassed the milestone of $1 billion dollars in funding for patient–centered outcomes research (PCOR) and comparative effectiveness research (CER). The goal of this investment is to conduct research that will improve decision making by patients, clinicians, payers and other stakeholders, resulting ultimately in improved public health and reduced health care spending [1]. One of the fundamental features distinguishing CER and PCOR from traditional clinical research is an emphasis on measuring and reporting outcomes that are more meaningful to patients, and that better reflect the decision making needs of clinicians, payers and policymakers [2, 3]. Systematic reviews of clinical trials have consistently observed problems with the outcomes reported in published studies, not only with respect to the relevance of those outcomes, but significant variation in which outcomes are reported, the instruments used to report them and biases from reporting some, but not all, of the outcomes that were collected in the trials. Because of this, the ability to use clinical trials to make reliable comparisons of the effectiveness of therapies is limited.

In this paper we provide an overview of the current state of outcomes reporting in clinical trials, describe a number of specific initiatives that are working to improve outcome reporting in future trials. The paper focuses most heavily on the increasing activity in the development of Core Outcomes Sets (COS), and the potential for this work to substantially improve the quality and relevance of outcomes measured and reported in clinical research. Finally, the paper identifies a number of gaps in this work and proposes a series of activities necessary to accelerate the development and use of more relevant, consistent and patient-centered outcomes. We believe that meaningful progress toward the goals behind the significant investments in PCOR and CER will depend on sustained attention to these challenges and proposed solutions.

Current problems with the outcomes in clinical trials

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. Currently there are three major problems with outcomes reported in clinical trials: 1) failure to collect outcomes that are most meaningful to patients, 2) a high degree of variability across trials in the outcomes reported, and 3) biased reporting of outcomes in published trials.

A number of reports have observed that the outcomes included in clinical research have not always been those that patients regard as most important or relevant [4]. For example, clinical guidelines issued by the American College of Physicians noted that the ability to provide strong recommendations was limited by the absence of measures of cognitive function that are commonly used in clinical care [5]. Medicare declined to provide reimbursement for cervical artificial discs in part because they viewed the outcomes reported in the trials as uninformative for key aspects of patient functional abilities. The primary outcome reported for most trials of drugs for psoriasis, which is also the one that is required for regulatory approval in the US, is based on a clinician's judgment of the extent of disease, while patients view the distribution of the plaques as most meaningful for their quality of life. In their 2012 methodology committee report, PCORI provides a standard for patient outcomes that instructs researchers to, "measure outcomes that people representing the population of interest notice and care about", to be identified with input from patients and decision makers through meetings, surveys or published studies [6]. While this standard should

raise awareness of the importance of patient-relevant outcomes, relying on individual research teams to select these outcomes independently and informally is unlikely to reduce the problems described further below.

The second major problem with outcomes in trials is variability in measurement and reporting. Evidence-based decision making often depends on the ability to aggregate results from multiple studies or to make treatment comparisons indirectly by looking at results of separate studies. These efforts depend heavily on the degree of consistency in the outcomes that are measured and reported across studies. At present, many studies which explore the effects of the same intervention on a specific health condition measure or report different outcomes, whether they be patient-reported or clinician-reported. This makes it difficult to compare, contrast or combine the findings of these studies when making decisions and setting policies, causing problems for people trying to use the output from healthcare research. For example, a survey of trials involving people with schizophrenia found that 2194 different scales had been used in 10,000 controlled trials: on average, a new instrument had been introduced for every fifth trial [7]. In other research, it has been shown that more than 25,000 outcomes appeared only once or twice in oncology trials [8].

Difficulties caused by heterogeneity in outcome measurement are well known to systematic reviewers [9] and hamper efforts to present guideline developers with succinct information on the most important outcomes. For example, Summary of Findings (SoF) tables were developed by the Grading of Recommendations Assessment, Development and Evaluation (GRADE) working group, to provide a summary of the evidence for important outcomes, along with the quality of this evidence [10]. They allow for the inclusion of up to seven important reported outcomes, providing a way to present the main findings of a review in a simple and transparent format. They have been shown to improve readers' understanding and speed of retrieval of the findings of systematic reviews [11] , but they will only be effective for decision-making if they include the most relevant outcomes for that purpose. However, a recent review found that although there has been an increased inclusion of SoF tables in Cochrane Reviews since they were introduced in 2008, they were still absent from nearly half of the reviews published for the first time in 2013 [12] .

The third major problem with outcomes in clinical trials is reporting bias. Outcome reporting bias (ORB) occurs as a consequence of the selection for publication of a subset of the originally collected outcomes on the basis of the results. This form of bias has been identified as a threat to evidence-based medicine because clinical trial outcomes with statistically significant results are more likely to be published [13]. The current CONSORT (Consolidating Standards of Reporting Trials) statement for reporting trials recommends that completely defined primary and secondary outcome measures should be pre-specified and any changes to trial outcomes after the trial commenced should be documented with reasons [14]. It goes on to recommend that the results for each outcome should be reported for each group, along with the estimated effect size and its precision. Despite this guidance, empirical research has shown that statistically significant outcomes were more likely to be fully reported compared with non-significant outcomes (range of odds ratios: 2.2 to 4.7). When comparing trial publications with protocols, it was found that 40 to 62% of studies had at least one primary outcome that was changed, introduced, or omitted in the time period between the production of these documents describing what the researchers planned to do and what they eventually did [15].

Definition and Examples of Core Outcome Sets

These issues of relevance, inconsistency and outcome reporting bias could be improved with the development and application of agreed standardised sets of outcomes, known as core outcome sets (COS), to be measured and reported for specific areas of health [16]. The outcomes included in a COS could include patient-reported outcomes, clinician-reported outcomes, and other patient-relevant outcomes. These sets are intended to represent the minimum that should be measured and reported in all clinical trials of a specific condition.

The existence or use of a COS does not imply that outcomes in a particular trial should be restricted to those in the relevant set. But, the expectation is that the core outcomes will always be collected and reported, with researchers including additional outcomes of particular relevance or interest to their specific study if they wish. The use of COS will make it easier for the results of trials to be compared, contrasted and combined, thereby reducing waste in research [17]. Their use would greatly reduce heterogeneity between trials because all trials

would measure and report the agreed important outcomes, and lead to research that is more likely to have measured patient-relevant outcomes. Importantly, their use would enhance the value of evidence synthesis by reducing the risk of outcome reporting bias and ensuring that all trials contribute usable information to a review and meta-analysis.

COS serve an analogous role to having a defined set of quality measures that are measured and reported consistently for all public reporting and pay for performance programs, such as those developed through the National Quality Forum. These measures have the benefit of being developed through a robust, transparent, multi-stakeholder process to ensure that they are relevant, reliable and efficient. They also allow for accurate comparisons across providers, and can be aggregated across multiple providers to generate meaningful information on trends in outcomes across larger health systems and regions.

An early example of an attempt to standardize outcomes was an initiative by the World Health Organization in the 1970s, relating to reporting results of cancer treatment trials [18]. More than 30 representatives from groups undertaking trials in cancer came together, the result of which was a WHO handbook recommending the minimum requirements for data collection in cancer trials. This data set included the minimum data that should be made available about the patient, the tumor, toxicity and effects of therapy including response, recurrence and disease-free survival.

Since then, particularly notable work relating to outcome standardization has been undertaken by the OMERACT (Outcome Measures in Rheumatology) initiative (www.omeract.org), which has advocated the use of COS, designed using consensus techniques, in clinical trials in rheumatology since their first conference in 1992 [19]. 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. As an example, the COS recommended for trials of medicinal products in the treatment of rheumatoid arthritis included the following seven outcomes: tender joints, swollen joints, pain, physician global assessment, patient global assessment, physical disability and acute phase reactants. There are currently 20 groups working on separate COS within the field of rheumatology, all coordinated under the umbrella of OMERACT.

The first evaluation of the uptake of a COS related to recommendations made for clinical trials of symptom-modifying anti-rheumatic drugs (SMARDS) in the treatment of rheumatoid arthritis (RA), ratified in 1994 by the WHO and International League of Associations for Rheumatology (ILAR), and also included in guidance issued by the FDA and EMA. This study demonstrated that nearly 70% of trialists reporting trials in RA are now measuring the COS [20], and 90% of the trialists contacted said they would consider using the COS if they were to lead a new trial in RA. Clearly, COS have the potential to improve the evidence base for health care, but additional work is need to develop strategies to ensure that they are disseminated and used by clinical researchers.

Since OMERACT, there have been other examples of similar COS initiatives to develop recommendations about the outcomes that should be measured in clinical trials. These include 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. Including the first IMMPACT meeting in 2002, there have been 17 consensus meetings on clinical trials of treatments for acute and chronic pain in adults and children. Additional examples are the Harmonizing Outcome Measures for Eczema (HOME, www.nottingham.ac.uk/homeforeczema) Initiative and the International Dermatology Outcomes Measures (IDEOM, www.dermoutcomes.org), which are international groups developing core outcomes to include in trials of the treatment of skin conditions.

The COMET Initiative

Stimulated in part by the success of OMERACT and IMMPACT, as well as the increase in awareness of problems with outcomes collected and reported in clinical trials, interest and activity in COS has been increasing rapidly over the past 5 years. The COMET (Core Outcomes Measures for Effectiveness Trials) Initiative (www.comet-initiative.org) was established to encourage and support the process of developing and implementing COS [21]. It was launched in 2010 with the following aims: 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, and 6) to encourage evidence-based COS development.

COMET aims to collate and stimulate the development, application and promotion of COS, by including data on relevant individual studies in a publically available internet-based resource. This database includes publications of previous COS development projects, as well as planned and ongoing work. A systematic review was undertaken in 2013 that provided a first comprehensive search for COS in health research [22]. It identified 198 relevant studies that determined which outcomes or domains should be measured in clinical trials for a specific health condition. The review revealed wide variation in the methods used to develop COS and work is needed to assess the implications of these different methods for both minimizing bias and maximizing efficiency in the development of COS, and for ensuring uptake. As an example, although benefits have been shown for involving patients in trial design, only 16% of the published COS reported that there was input from patients in their development [23]. The review highlighted the need for methodological guidance, including how to engage key stakeholder groups, particularly members of the public, in the development and implementation of COS.

This review was updated to the end of 2014, and a further 29 new COS studies were identified [24]. There has been a general increase in the number of COS over the years, with a consistently higher number of COS published annually in recent years than in most years before 2010 (Figure 1). The studies identified have been added to the COMET database in order to provide an up-to-date, comprehensive database of COS. In addition, to reduce unnecessary duplication of effort, ongoing COS studies are also registered in the database. Figure 2 shows the number of COS developed, or in development, according to health category. Taken together, this work has identified many health areas where a COS has been developed and highlighted important gaps. For example, there has been substantial work on COS for rheumatology and neurology, while mental health conditions have not received significant attention.

As awareness of the need for COS continues to grow and knowledge of the COMET Initiative increases, this is reflected in the website and database usage figures [17, 25]. More than 16,500 visits were made to the website in 2014 (36% increase over 2013) and 9780 new visitors (43% increase). By December 2014, a

total of 6588 searches had been completed of the COMET database, with 2383 in that year alone (11% increase). An online survey in May-June 2015 was answered by 206 (52%) of the 396 visitors to the website, and revealed that the most common reasons for searches were 'I am thinking about developing a core outcome set' and 'I am planning a clinical trial'.

Methods for Developing COS

While standardized methods for the development of COS have not yet been widely adopted, a fairly well-defined set of issues to consider are increasingly being addressed [16]. A detailed handbook with step-by-step instructions on the COS development process has been developed by OMERACT, and is regularly updated with insights generated through ongoing COS work [26]. At a high level, the first stage in the development of a COS is most frequently a decision on what outcomes or outcome domains to measure, followed by agreement on how those outcomes should be defined and measured in order to provide the necessary data, adopting the outcome specification model previously described [27]. COS may include both patient-reported outcomes (PROs) and clinician-reported outcomes (ClinROs). Several existing initiatives, described in the next section, are relevant to the process of determining how an outcome should be defined and measured; and a practical guideline has recently been developed [28].

As noted above, COS have been developed using a variety of methods [22, 24]. These include the use of a literature/systematic review as an early step, which rose from 33% (66/198) of the studies in the original review to 72% (21/29) in the update. A variety of methods have been used to assess and develop consensus, with the Delphi technique used for 15% (29/198) of the COS in the original review and 31% (9/29) in the update.

The stakeholder groups regarded as key to developing a COS varies between health areas. Clinical experts have been involved in all studies, but there has been a recent shift towards greater involvement of patient and public representatives. This increased, where reported, from 18% (31/174) in the original review, to 59% (13/22) in the update, and 89% (66/74) among the ongoing studies that are registered in the COMET database.

Involving the public in research can bring challenges and, in response to this need, COMET has established the PoPPIE (People and Patient Participation, Involvement and Engagement) Working Group to develop resources, including the development of plain language summaries in partnership with patients (www.comet-initiative.org/resources/PlainLanguageSummary), and pursue a research agenda. Research is needed on how to: involve patients as research partners in design of COS studies; identify and meet information needs of patients as both research partners and participants; identify appropriate methods for eliciting consensus among patient groups; generate appropriate questions for patients taking part in a COS study; access and engage patients in COS studies; ensure hard to reach communities are involved; bring different stakeholder groups' views together; and evaluate the stakeholder experience of taking part.

Other initiatives involving standardized outcomes

A number of programs and initiatives in the US and elsewhere overlap to some degree with the expanding work on COS, and present opportunities for coordination and collaboration, as well as the need to avoid duplication of effort.

PROMIS (www.nihpromis.org) is an NIH initiative which provides a system of measures of patient-reported health status for physical, mental, and social health and can be used in studies across a range of medical conditions. A primary objective of PROMIS is to assemble a set of questions to assess the most common dimensions of patient–reported outcomes for a wide range of chronic diseases. These include items to measure pain, fatigue, psychological distress, physical function, and overall health. The PROMIS database of measures focusses on patient-reported outcomes, while COS generally include both patient-reported and clinician-reported outcomes, as well as other outcomes such as laboratory results, etc. PROMIS focusses on 'how' to measure 'items' that are patient-reported, providing an item question bank. The 'how' is typically the second stage of a COS, with the 'what' coming first. In addition, PROMIS does not seek to develop agreement around a defined set of minimum outcomes to measure and report in a standard fashion across studies, and, by itself therefore, would not address the fundamental problems of relevance, consistency or reporting bias discussed above.

The National Institutes of Health (NIH) has also taken steps to encourage the use of common data elements (CDEs), including outcomes, in NIH supported research projects or registries. The NIH provide a resource portal (www.nlm.nih.gov/cde) that includes databases and repositories of data elements and case report forms that might help investigators in identifying and selecting data elements for use in their projects. The NIH CDE program is most relevant to researchers once they have determined 'what' to measure, offering resources on 'how' to measure both particular outcomes and other relevant data items.

The International Consortium for Health Outcomes Measurement (ICHOM, www.ichom.org) organizes global teams of physician leaders, outcomes researchers and patient advocates to define core sets of outcomes for use for specific health conditions to assess the quality of clinical practice. This initiative was established with the aim of providing a structured process to achieve consensus data collected standards to be used outside the context of clinical trials, and hopes to publish 50 standard sets by 2017. A list of completed sets, in progress and conditions under consideration is available at www.ichom.org/medical-conditions.

ICHOM is focusing on the development of core data sets to evaluate the quality and efficiency of clinical care, rather than for use in clinical trials. The methods used by ICHOM differ substantially from those used in the development of COS for clinical trials (even taking account of the variability in methods used for the latter), as discussed in more detail below. It is unclear whether the outcomes sets developed by ICHOM for clinical care would also be useful for clinical research, although there would be value in developing standards that could be used for both purposes, making it possible to re-use data collected for either reason.

As an example, ICHOM recently published their consensus measures for patients treated for prostate cancer [29], but provided limited detail on the specific methods used to reach agreement on the recommended measures. It would be valuable to have more information on those methods, such as how the number of patients and other stakeholders was decided; how patients were identified, selected and involved in the process; how final decisions of the inclusion of an outcome were made [30]. In addition, it is important to know how the measuring tools were selected. For instance, the ICHOM standards require a record of the date of recurrence of an abnormal PSA , but the variety of definitions of PSA recurrence (either within a treatment or across treatments) could render comparisons difficult and problematic, particularly for clinical trials [30].

The FDA Clinical Outcome Assessments (COA) Staff aim to encourage the development and application of patient-focused endpoint measures in medical product development to highlight clinical benefit in labeling. They 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. Unmet medical needs are addressed through the Clinical Outcomes Assessment Qualification Program. COA qualification is dependent on appraisal of the evidence to support the conclusion that the COA is a well-defined and reliable assessment of a particular concept of interest for application in studies used to support drug marketing authorization. The European Medicines Agency (EMA) is working along similar lines to increase the use of well validated patient-relevant outcomes in their regulatory process.

The main focus of the FDA and EMA work is understandably on the endpoints to be used in product approvals and labelling, and has not been concerned to date with the assessment of effectiveness involving outcomes that may be useful for non-regulatory decisions by patients, clinicians, payers and others. From a regulatory perspective, the emphasis is on outcomes that are customized to specific products and indications, rather than aiming for greater standardization of outcomes across products and studies. Furthermore, the degree of focus on patient-relevant outcomes, and inclusion of patients and the public has been limited to date. This may be changing as a result of greater attention to "patient-focused drug development", and may ultimately lead the FDA to greater consideration of outcomes that are relevant to and informed by patients [31]. For example work on Core Symptom Measures for cancer trials has been promoted in prostate, head and neck and ovarian cancers [32]. Whilst the processes to elicit expert opinion has included some patient involvement, it is quite limited [33] and greater emphasis of wider patient involvement in the process of selecting core outcomes is recommended.

In considering the measurement instruments to use, several may exist for any given outcome, usually with varying psychometric properties (i.e. reliability and validity). Systematic reviews of measurement instruments provide one way to select a measurement instrument for an outcome within a COS. The COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) initiative collates systematic reviews of studies of measurement properties of existing 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 at www.cosmin.nl/systematic-reviews-of-measurement-properties-_5_0.html.

Key challenges and moving forward

While it is encouraging that the level of activity around COS has been increasing, substantial work remains to be done, and there are a number of key challenges that must be addressed to accelerate progress. To inform how best to move forward, the Center for Medical Technology Policy and the COMET initiative hosted a one day COS workshop in April 2014, supported with funding from PCORI, The European Union and the UK Medical Research Council. The meeting opened with presentations by ten North American experts currently developing COS for use in clinical trials, patient-centered outcomes research (PCOR), systematic reviews, quality improvement and other contexts. Each presentation was followed by discussion and feedback from representatives of federal agencies and national organizations with an interest in condition-specific, standardized health outcomes, including most of those described above. The discussions underscored the need to expand capacity to develop high quality COS, and identified several issues requiring attention to promote further progress. These challenges and potential strategies to address them are summarized below.

Better understanding of the gaps in COS

The number and quality of COS remains limited, despite the recent increase in activity in this field. While a systematic review of gaps has not yet been done, an initial attempt has been made to map the content of the COS database to the most prevalent acute and chronic conditions. For example, no COS have been published regarding trials of drug therapy for type 2 diabetes or interventions for the management of chronic wounds. And for many other conditions, the available COS have used informal consensus methods rather than structured approaches, which might undermine their acceptability for a wide range of decision makers. An initial informal review of conditions identified by the WHO as being responsible for the highest global burden of disease [34] has identified key gaps. A more systematic assessment of these gaps is now underway. Again, even if existing COS have been identified, further work is needed to determine whether they are of adequate quality to broadly promote.

An important next step would be to conduct a systematic assessment to identify those high prevalence, high burden conditions for which high quality COS do not yet exist. This could be done by using work that has already been done to identify and rank burden of illness globally or nationally. For example, in the US, the Agency for Health Care Research and Quality has developed priorities for comparative effectiveness research, as have the Institute of Medicine and PCORI. Having identified the priority conditions, published COS in those clinical domains could be found relatively simply through the COMET database. Finally, these COS would be need to be assessed for quality, of both their methods and reporting. As noted below, work is underway to develop quality assessment and reporting tools for COS, but the initial quality screening could look for a few basic indicators of quality, such as the inclusion of patients or consumers in the development process.

Expanded capacity to produce COS

As priority areas for COS are identified, mechanisms need to be in place to support their development and implementation. This will require research to identify best practices for COS development and the production of a reporting guideline to facilitate clear reporting of the COS and the processes used to develop it (see below). This would likely be accelerated by an increased interest in research funding agencies to support the development of high priority COS, and of the methods and tools necessary to support groups involved in this work. The nature of this work requires multi-stakeholder collaboration, ideally at the national or international level. Ideally, many of these initiatives would include leadership from the patient advocacy organizations relevant to each topic. In this way, the benefits of COS for reducing waste in research will not arise from wasteful practices in their development [35] .

As part of its effort to support COS developers, the COMET database includes previous work that might help the development of new COS, alongside the reports of COS themselves. For example, an initial step in COS development is usually a review of outcomes measured in previous clinical trials and work to identify outcomes felt to be important by patients. With this in mind, the database includes more than 120 reviews of outcomes measured in trials and 52 studies of patients' perspectives on outcomes to be measured in their condition.

Improving the quality of COS

The afore-mentioned systematic review and its update revealed wide variation in the methods used to develop COS, with no clear consensus on best practice ([22]; Gorst et al, in preparation). Although key issues to consider when developing a COS have been described [16], there is little or no guidance on how to choose and involve stakeholders, develop consensus, achieve geographical representation, or undertake many other aspects of the process.

One of the urgent areas of need is to expand engagement of patients and the public in the development of COS. As noted above, a significant minority of published reported any patient or public involvement in the process. There is anecdotal evidence to suggest that this has started to shift, and the forthcoming updated systematic review of COS will provide more recent empirical data. Given the rapidly increasing recognition of the need to develop, validate and report outcomes that are meaningful to patients, it is a critical priority to fund methods research to develop and validate formal qualitative methods to effectively engage patients and the public in this process. The absence of empirically-based best practices should not prevent the development of initial consensus around key principles and techniques. Once documented and standardized to some extent, it will be possible to do empirical work to evaluate the effectiveness of alternative approaches.

When best practices have been developed, it will also be possible and useful to develop a quality assessment tool. Although nearly 230 published COS studies have already been identified, the lack of an assessment tool means that there has been no formal quality assessment of these. Defining the quality of a COS is not straightforward. Ultimately, a "good" COS would be one that leads to improved health outcomes but this might be far down-stream of the development process and difficult to measure. It is also unlikely to be a feature of any report describing the development of a COS. Instead, a tool is needed to assess how the COS developers minimized biases which would otherwise undermine the ability of the COS to have a positive impact on patient care and outcomes. This would be analogous to the Cochrane Risk of Bias tool for assessing the studies to include in systematic reviews [36].

Along with best practice guidance and metrics to assess the quality of COS, it will be valuable to develop a standardized reporting tool for publications and reports of COS. Although a preliminary checklist was proposed for the reporting of Delphi surveys used in the development of COS [37], this is not sufficient to address the

wider aspects of COS studies. COS studies are still not well reported in terms of the scope and methods used in the development of the COS. Information that might help users to decide whether to adopt a COS or develop a new one is often lacking. To try to redress this, work is underway to develop a COS reporting guideline, using an international consensus process [38]. The initial areas under consideration include the rationale for and scope of the COS, methods of development, stakeholder involvement, sources of information, consensus process, limitations, and plans for implementation and updating.

Increasing uptake of high quality COS

Where high quality COS exist or are developed, they will need to be widely accepted and implemented by the majority of researchers if their benefits for health are to be realized. Publication and broad dissemination of the COS is a basic requirement, and the simplified access to relevant reports and papers is now possible as a result of the work of COMET. Stronger incentives are likely to be necessary for high levels of uptake, for example, the recognition or endorsement of COS by research funding organizations, journal editors, clinical guideline developers and payer/HTA organizations. Researchers should be considerably more likely to give serious consideration to COS when that decision has the potential to impact the funding of their research, publication of their results, of have an influence over clinical policy or reimbursement.

With respect to research funders, the National Institute for Health Research (NIHR) in the UK provides the following text in their guidance notes for applicants: "Where established Core Outcomes exist they should be included amongst the list of outcomes unless there is good reason to do otherwise. Please see The COMET Initiative website at www.comet-initiative.org to identify whether Core Outcomes have been established". There are not yet examples of US funders making explicit reference to COS, but the NIH, AHRQ, PCORI and others may wish to consider under what circumstances it would be reasonable to do so.

Some journal editors have begun to encourage consideration of COS by researchers. The SPIRIT guidance for reporting protocols of clinical trials [39] recommends "Where possible, the development and adoption of a common set of key trial outcomes within a specialty can help to deter selective reporting of outcomes and to facilitate comparisons and pooling of results across trials in a meta-analysis. The COMET (Core Outcome Measures in Effectiveness Trials)

Initiative aims to facilitate the development and application of such standardized sets of core outcomes for clinical trials of specific conditions (www.comet-initiative.org). Trial investigators are encouraged to ascertain whether there is a core outcome set relevant to their trial and, if so, to include those outcomes in their trial. Existence of a common set of outcomes does not preclude inclusion of additional relevant outcomes for a given trial". Finally, in obstetrics and gynecology, the CROWN (Core Outcomes in Women's Health) Initiative (www.crowninitiative.org) is a consortium of more than 70 journal editors that will "strongly encourage the reporting of results for core outcome sets. Facilitate embedding of core outcome sets in research practice, working closely with researchers, reviewers, funders and guideline makers".

There is considerable potential for increasing uptake of COS through the mechanism of recognition by clinical guideline developers, HTA organizations and payers – as these groups all carry substantial influence through their direct influence on the evidence-based reimbursement of products and services. The manual for guideline development by the National Institute for Health and Care Excellence (NICE) recommends the COMET database as a source of information to be considered (NICE, 2012). The Green Park Collaborative (GPC) in the US (www.cmtpnet.org/featured-projects/green-park-collaborative) develops recommendations for research in specific therapeutic areas, through a multi-stakeholder collaborative process that includes payers, guideline developers, HTA organizations, as well as patients, clinicians and other key stakeholders. A number of GPC projects have attempted to increase the measurement and reporting of patient-relevant outcomes [40]. Efforts are also underway to leverage the broad range of decision making authority represented by the membership of this collaborative to promote the uptake of high quality COS. Ultimately, it is likely that explicit recognition or formal endorsement of COS will be necessary to ensure the level of consistent use that will achieve the original objectives of this work.

Conclusions

If comparative effectiveness research is to have an impact on practice and, thereby, the health of patients and the public, it needs to ensure that the outcomes that matter most to the people making decisions and choices are measured and reported. Core outcome sets can facilitate this, especially when their development and implementation is integrated with other key initiatives to

improve research and practice. There are several challenges to overcome. Work is already underway to meet some of these, including a database of all identified COS, research to identify best practices and the development of a reporting guideline. However, much remains to be done. Health areas in particular need of COS have to be identified, the development of high quality COS to meet this need has to be encouraged and supported, and these COS have to be implemented in research and practice.

In addition to standardized outcomes in clinical trials, consistent reporting of outcomes is also important for studies of the onset, course and consequences of disease. To this end, clinical cohorts, registries and observational databases have been established; and routinely collected data from within electronic medical record (EMR) systems is likely to become increasingly available in the future to support research. The EMR provides real-world clinical information and is a sustainable source of data. The overlap between data from research databases and from EMR systems for clinical care could be valuable. Therefore, to achieve the full potential of such an approach, it is vital that similar attention is paid to the choice and definition of outcomes that should be collected within such systems. It has also been proposed that prospective registries of clinical trials should place increased emphasis on COS [41]. For example, registries could encourage researchers to note their use of a COS when they register their trial, perhaps with a prompt for them to consider using a COS and a simple way for them to specific the outcomes by a link through to the original COS.

Funding Support

This work was supported in part by a Eugene Washington grant from PCORI.

This work was also funded in part by the MRC MRP (Medical Research Council Methodology Research Panel), grant number MR/J004847/1; and European Union Seventh Framework Programme ([FP7/2007-2013] [FP7/2007-2011]) under grant agreement nu 305081.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The views expressed in this article do not necessarily represent the views of the funders.

References

1. Orszag PR, Ellis P. Addressing rising health care costs--a view from the Congressional Budget Office. *The New England journal of medicine* 357(19), 1885-1887 (2007).
2. Gabriel SE, Normand SL. Getting the methods right--the foundation of patient-centered outcomes research. *The New England journal of medicine* 367(9), 787-790 (2012).
3. Sox HC, Greenfield S. Comparative effectiveness research: a report from the Institute of Medicine. *Annals of internal medicine* 151(3), 203-205 (2009).
4. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet* 374(9683), 86-89 (2009).
5. Qaseem A, Snow V, Cross JT, Jr. *et al*. Current pharmacologic treatment of dementia: a clinical practice guideline from the American College of Physicians and the American Academy of Family Physicians. *Annals of internal medicine* 148(5), 370-378 (2008).
6. Patient Centered Outcomes Research Institute Methodology Committee. The PCORI methodology report. 2015(25 Aug), (2013).
7. Miyar J, Adams CE. Content and quality of 10,000 controlled trials in schizophrenia over 60 years. *Schizophrenia bulletin* 39(1), 226-229 (2013).
8. Hirsch BR, Califf RM, Cheng SK *et al*. Characteristics of oncology clinical trials: insights from a systematic analysis of ClinicalTrials.gov. *JAMA internal medicine* 173(11), 972-979 (2013).
9. Smith V, Clarke M, Williamson P, Gargon E. Survey of new 2007 and 2011 Cochrane reviews found 37% of prespecified outcomes not reported. *J Clin Epidemiol* doi:10.1016/j.jclinepi.2014.09.022 (2014).
10. Guyatt G, Oxman AD, Akl EA *et al*. GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 64(4), 383-394 (2011).
11. Rosenbaum SE, Glenton C, Oxman AD. Summary-of-findings tables in Cochrane reviews improved understanding and rapid retrieval of key information. *J Clin Epidemiol* 63(6), 620-626 (2010).
12. Wuytack F, Smith V, Clarke M, Williamson P, Gargon E. Towards core outcome set (COS) development: a follow-up descriptive survey of outcomes in Cochrane reviews. *Systematic reviews* 4 73 (2015).
13. Kirkham JJ, Dwan KM, Altman DG *et al*. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 340 c365 (2010).
14. Schulz KF, Altman DG, Moher D, Group C. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *PLoS Med* 7(3), e1000251 (2010).
15. Dwan K, Gamble C, Williamson PR, Kirkham JJ, Reporting Bias G. Systematic review of the empirical evidence of study publication bias and outcome reporting bias - an updated review. *PLoS ONE* 8(7), e66844 (2013).
16. Williamson PR, Altman DG, Blazeby JM *et al*. Developing core outcome sets for clinical trials: issues to consider. *Trials* 13(1), 132 (2012).
17. Gargon E, Williamson PR, Altman DG, Blazeby JM, Clarke M. The COMET Initiative database: progress and activities from 2011 to 2013. *Trials* 15(1), 279 (2014).
18. Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. *Cancer* 47(1), 207-214 (1981).
19. Tugwell P, Boers M, Brooks P, Simon L, Strand V, Idzerda L. OMERACT: an international initiative to improve outcome measurement in rheumatology. *Trials* 8 38 (2007).
20. Kirkham JJ, Gargon E, Clarke M, Williamson PR. 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 (2013).

21. Williamson P, Altman D, Blazeby J, Clarke M, Gargon E. Driving up the quality and relevance of research through the use of agreed core outcomes. *Journal of health services research & policy* 17(1), 1-2 (2012).
22. Gargon E, Gurung B, Medley N *et al*. Choosing important health outcomes for comparative effectiveness research: a systematic review. *PLoS ONE* 9(6), e99111 (2014).
23. Staley K. Exploring impact: public involvement in NHS, public health and social care research. (2009).
24. Gorst SL, Gargon E, Clarke M, Blazeby JM, Altman DG, Williamson PR. Choosing important health outcomes for comparative effectiveness research: an updated review and user survey. *PLoS ONE, submitted* (2015).
25. Gargon E, Williamson PR, Altman DG, Blazeby JM, Clarke M. The COMET Initiative database: progress and activities update (2014). *Trials, submitted* (2015).
26. Maarten Boers, Kirwan JR, Tugwell P *et al*. The OMERACT Handbook. (2015).
27. Zarin DA, Tse T, Williams RJ, Califf RM, Ide NC. The ClinicalTrials.gov results database--update and key issues. *The New England journal of medicine* 364(9), 852-860 (2011).
28. Prinsen CA, Vohra S, Rose MR *et al*. How to select outcome measurement instruments for outcomes included in a 'Core Outcome Set' – a practical guideline. *Trials, submitted* (2015).
29. Martin NE, Massey L, Stowell C *et al*. Defining a standard set of patient-centered outcomes for men with localized prostate cancer. *European urology* 67(3), 460-467 (2015).
30. Maclennan S, Williamson PR, Lam TB. 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. *European urology* doi:http://dx.doi.org/10.1016/j.eururo.2015.08.015 (2015).
31. Basch E. Toward patient-centered drug development in oncology. *The New England journal of medicine* 369(5), 397-400 (2013).
32. Reeve BB, Mitchell SA, Dueck AC *et al*. Recommended patient-reported core set of symptoms to measure in adult cancer treatment trials. *J Natl Cancer Inst* 106(7), (2014).
33. Main BG, Blencowe N, Williamson PR, Blazeby JM. RE: Recommended patient-reported core set of symptoms to measure in adult cancer treatment trials. *J Natl Cancer Inst* 107(4), (2015).
34. Global Burden of Disease Study Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* doi:10.1016/s0140-6736(15)60692-4 (2015).
35. Macleod MR, Michie S, Roberts I *et al*. Biomedical research: increasing value, reducing waste. *The Lancet* 383(9912), 101-104 (2014).
36. Higgins JP, Altman DG, Gotzsche PC *et al*. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 343 d5928 (2011).
37. Sinha I, Smyth RL, Williamson PR. 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* 8(1), e1000393 (2011).
38. Kirkham JJ, Gorst S, Altman DG *et al*. COS-STAR: a reporting guideline for studies developing core outcome sets (protocol). *Trials* 16(1), 373 (2015).
39. Chan AW, Tetzlaff JM, Gotzsche PC *et al*. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. *BMJ* 346 e7586 (2013).
40. Basch E, Abernethy AP, Mullins CD *et al*. Recommendations for incorporating patient-reported outcomes into clinical comparative effectiveness research in adult oncology. *J Clin Oncol* 30(34), 4249-4255 (2012).
41. Clarke M, Williamson P. Core outcome sets and trial registries. *Trials* 16 216 (2015).

Figure 1 Year of first publication of each COS study (n=227)

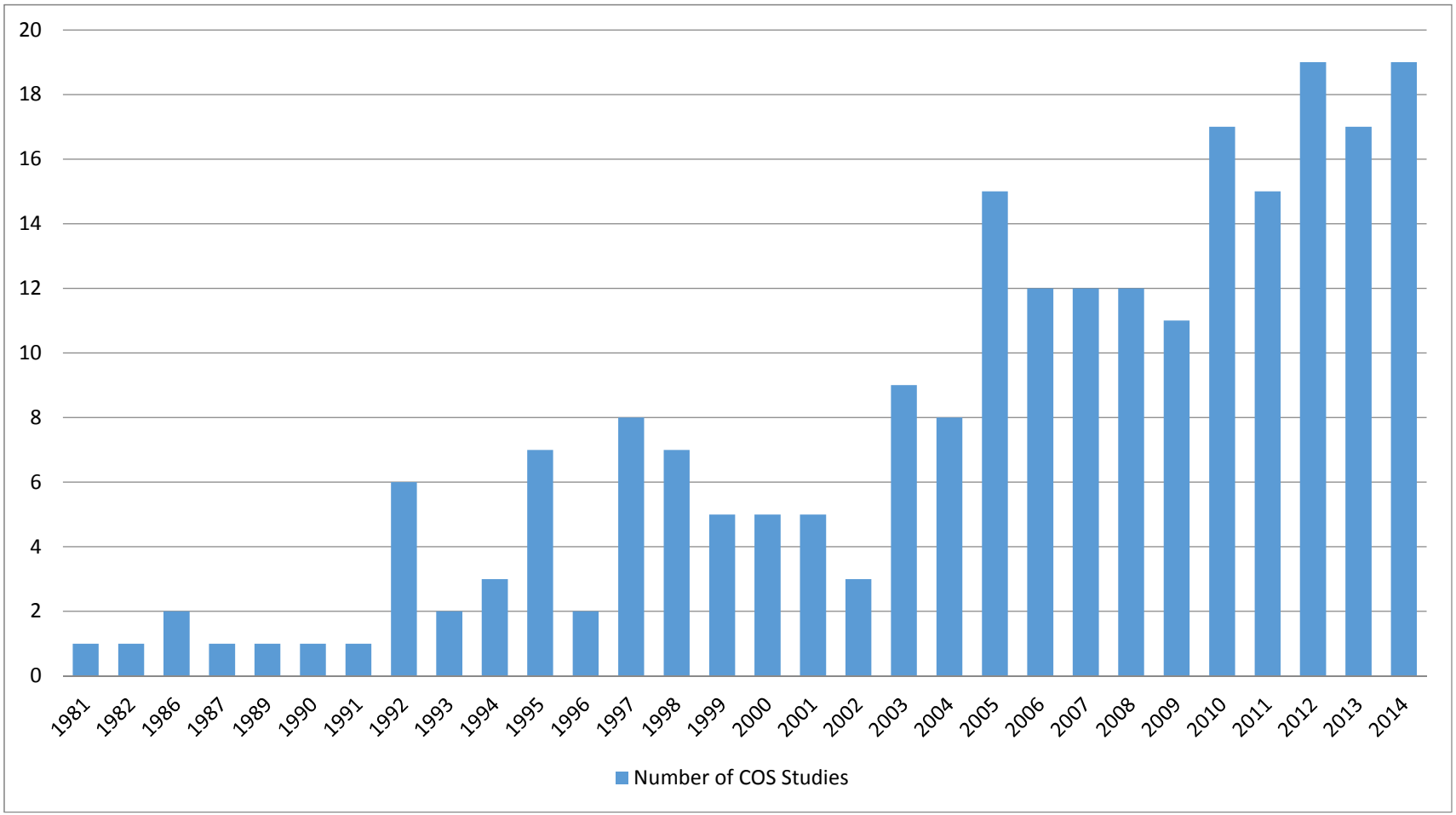
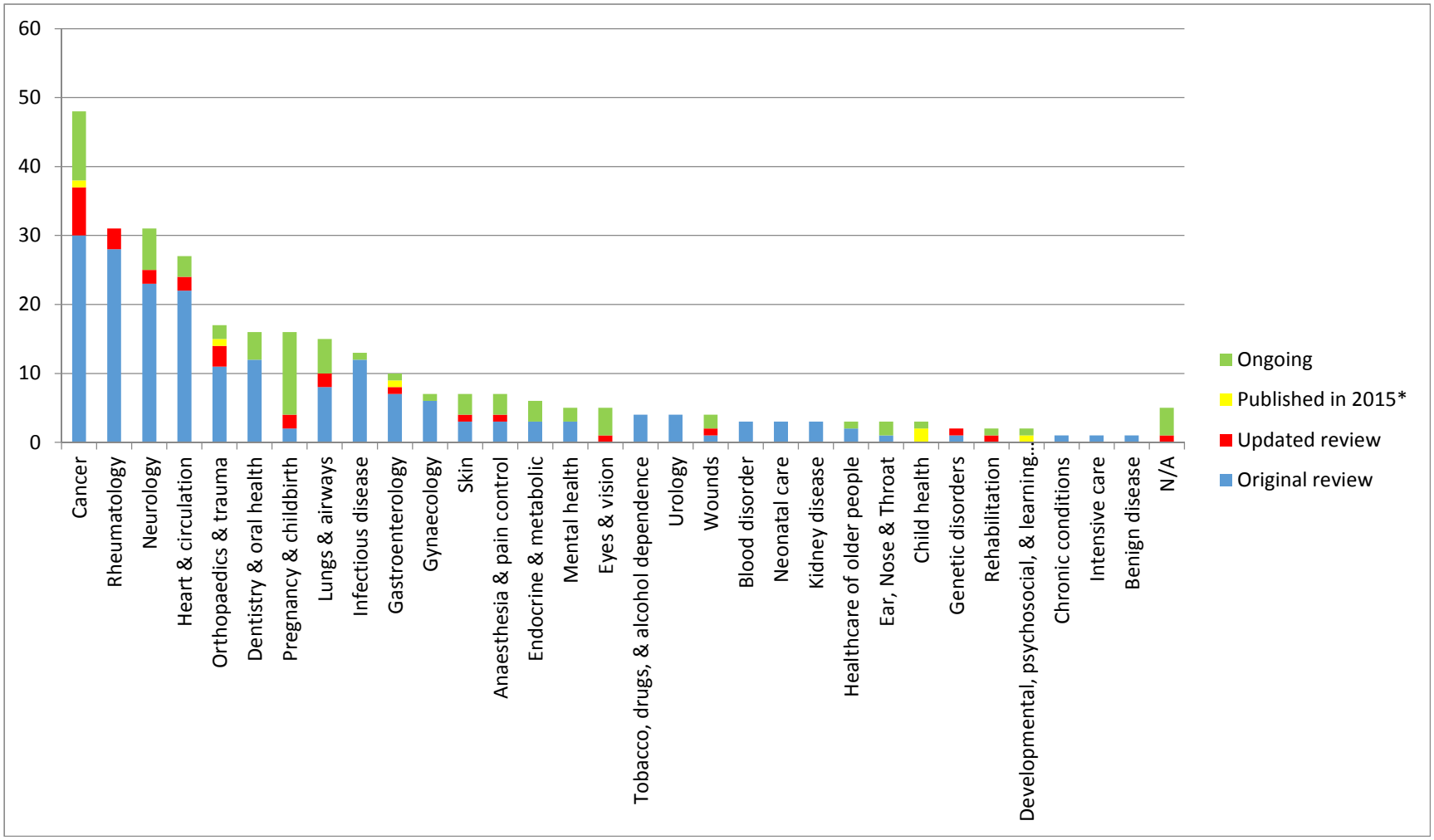


Figure 2 Number of COS developed in each disease category (n=305)



**Studies we are aware have been published since December 2014*