

A Systematic Review of Studies That Aim to Determine Which Outcomes to Measure in Clinical Trials in Children

Ian Sinha^{1*}, Leanne Jones¹, Rosalind L. Smyth¹, Paula R. Williamson²

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

Funding: Ian Sinha is funded by the Department of Health. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Academic Editor: David Moher, Children's Hospital of Eastern Ontario Research Institute, Canada

Citation: Sinha I, Jones L, Smyth RL, Williamson PR (2008) A systematic review of studies that aim to determine which outcomes to measure in clinical trials in children. *PLoS Med* 5(4): e96. doi:10.1371/journal.pmed.0050096

Received: November 9, 2007

Accepted: March 14, 2008

Published: April 29, 2008

Copyright: © 2008 Sinha et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abbreviations: CSG, Clinical Study Group; EMEA, European Medicines Agency; FDA, Food and Drug Administration; FEV1, forced expiratory volume in one second; GVHD graft-versus-host disease; IIM, idiopathic inflammatory myopathy; IMACS, International Myositis Assessment and Clinical Studies group; MCRN, Medicines for Children Research Network; NDDI, Neonatal Drug Development Initiative; NGT, nominal group technique; PIP, Paediatric Investigation Plan; PRINTO, Paediatric Rheumatology International Trials Organisation; SLE, systemic lupus erythematosus

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

ABSTRACT

Background

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

Methods and Findings

We searched Cochrane databases (no date restrictions) in December 2006; and MEDLINE (1950 to 2006), CINAHL (1982 to 2006), and SCOPUS (1966 to 2006) in January 2007 for studies of the selection of outcomes for use in clinical trials in children. We also asked a group of experts in paediatric clinical research to refer us to any other relevant studies. From these articles we extracted data on the clinical condition of interest, description of the method used to select outcomes, the people involved in the selection process, the outcomes selected, and limitations of the method as defined by the authors. The literature search identified 8,889 potentially relevant abstracts. Of these, 70 were retrieved, and 25 were included in the review. These studies described the work of 13 collaborations representing various paediatric specialties including critical care, gastroenterology, haematology, psychiatry, neurology, respiratory paediatrics, rheumatology, neonatal medicine, and dentistry. Two groups utilised the Delphi technique, one used the nominal group technique, and one used both methods to reach a consensus about which outcomes should be measured in clinical trials. Other groups used semistructured discussion, and one group used a questionnaire-based survey. The collaborations involved clinical experts, research experts, and industry representatives. Three groups involved parents of children affected by the particular condition.

Conclusions

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

The Editors' Summary of this article follows the references.

Introduction

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

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

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

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

Outcomes can reflect various effects of an intervention. They may directly measure a definitive clinical change, such as death or hospital admission. Surrogate outcomes, which are sometimes used in lieu of a definitive clinical outcome, aim to capture the effects of an intervention without having

to wait for the clinical change to actually occur. In other words, they are proximal to the clinical outcome on the disease pathway, so a change can be detected sooner. They may be a measure of intermediate health status, which may be used to predict future health status; for example, glycosylated haemoglobin is used as a measure of current disease control in patients with diabetes mellitus, and has been shown to be a useful predictor of future control [3]. A surrogate outcome may even be an assumed or established risk factor that actually impacts on disease progression; for example, neonatal intraventricular haemorrhage, which is a recognised complication of prematurity, is thought to alter brain development in the early stages of life and predispose babies to developmental problems in childhood. There are validation criteria that should be fulfilled before a surrogate outcome can be confidently used in place of a definitive clinical outcome in clinical trials [4].

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

Selecting Outcomes for Use in Clinical Trials

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

Table 1. Definitions Used in Review

Terminology	Description	Examples
Outcome domain	A relatively broad aspect of the effect of illness on a child, within which an improvement may occur in response to an intervention. In general these domains may not be directly measurable themselves, so outcomes are selected to assess change within them.	(1) In clinical trials of children with diabetes, outcome domains may include acute metabolic complications, long-term glucose control, renal damage, or effects on schooling. (2) In clinical trials of children with asthma, outcome domains may include lung function, health care utilisation, and symptom control
Outcome/endpoint	A measurable variable within an outcome domain. The outcome can be measured at a variety of time points, which must be clearly stated by authors of clinical trials.	(1) Absolute FEV1 expressed as change from baseline (2) Number of admissions to hospital within a six month period (3) Time to first seizure after starting an antiepileptic intervention
Composite outcome	An outcome that encompasses more than one domain. A composite outcome may span more than one outcome domain. These are sometimes referred to as global outcomes, because they theoretically measure several aspects of the impact of illness on a patient.	In the NIMH Collaborative Multisite Multimodal Treatment Study of Children with Attention Deficit/Hyperactivity Disorder, response to treatment was measured by a composite outcome comprising outcomes that reflect clinical outcome, academic functioning, effects on family life, and other domains [37]
Outcome measure	A scale, scoring system, questionnaire, or other tool used for measuring an outcome. They may be an amalgamation of more than one outcome within an outcome domain, such as a score based on a variety of symptoms. More than one outcome measure may be possible to use to represent change in an outcome.	(1) The Paediatric Asthma Quality of Life Questionnaire [38] (2) Childhood Health Assessment Questionnaire [39] (3) Gastro-esophageal Reflux Disease Symptom Questionnaire [40]

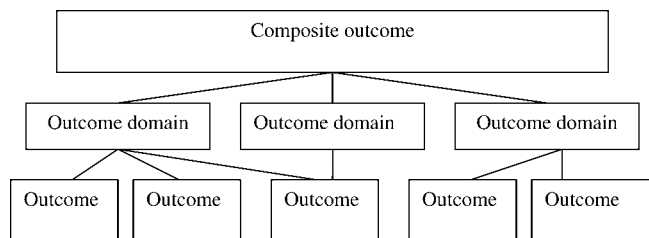


Figure 1. Levels of Outcome Measurement
doi:10.1371/journal.pmed.0050096.g001

tion. Examples of these problems are well documented [6,7]. Investigators can select from a range of several potential outcomes spanning different domains when they are designing a clinical trial, however, and the process of determining which outcome to use can be complex. The difficulty of selecting the most appropriate outcomes for use in a clinical trial is reflected in the fact that in several fields of clinical research there is much heterogeneity between clinical trials of specific diseases regarding exactly which outcomes to select [8,9]. Some of the factors underlying this uncertainty may be that for these conditions there is uncertainty about which outcome domains are most relevant to patients, that the performance characteristics of potential outcomes have not been established, or that as the general care of patients has improved, previously used outcomes are no longer relevant.

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

The objective of this project was to systematically review studies that address the process of selecting which outcome domains or outcomes to measure in clinical trials in children under 16 years of age. We have restricted this review to studies in children for the following reasons. In clinical research there is increasing recognition that children are not merely “small adults,” and the methodology of conducting research in a paediatric population should be tailored accordingly. We anticipate that one way of determining which outcomes to use, in addition to the consensus techniques described above, may be to ascertain the opinions of both children and their parents regarding what they think are important aspects of their disease. This process poses unique challenges that may not be relevant when conducting similar research in adults with a particular condition, so it is appropriate to specifically review studies pertaining to outcome domain and outcome selection in clinical trials in children.

Methods

Study Selection

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

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

Identification of Studies

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

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

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

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

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

Data Extraction

The following data were extracted by one reviewer (IS) and checked independently by the second reviewer (LJ): (1)

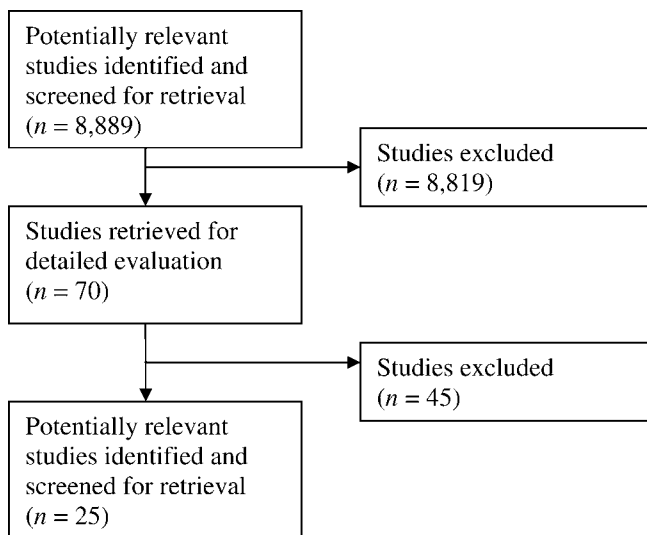


Figure 2. Study Flowchart
doi:10.1371/journal.pmed.0050096.g002

Condition for which the outcome domains or outcomes are discussed; (2) Description of the method; (3) People involved in selecting outcome domains or outcomes; (4) Outcome domains or outcomes selected; (5) The geographical setting of the collaborations, ascertained either by reading the text or, where listed, the names and institutions of people involved in the collaborations; (6) Limitations of the method as defined by the authors

Assessment of Methodological Quality

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

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

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

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

Data Analysis and Presentation of Results

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

Results

Description of Studies

The initial database search identified 8,889 potentially relevant abstracts, of which 70 articles were retrieved in full

and, finally, 25 included in the full review, as depicted by the flowchart in Figure 2.

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

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

Agreement between Reviewers

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

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

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

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

Included Studies

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

Six of these groups (Griffiths et al. [11], Ramsey et al. [12], Pavletic et al. [13], Giannini et al. [14], International Myositis Assessment and Clinical Studies group (IMACS) [15], and Paediatric Rheumatology International Trials Organisation (PRINTO) [16]) aimed to develop a consensus statement specifically about outcome measures that should be used in clinical trials of certain medical conditions. Five groups (Carlson et al. [17], Goldstein et al. [18], LaFrance et al. [19], Neonatal Drug Development Initiative (NDDI) [20–24], and West Delphi group [25]) discussed which outcomes to measure as part of workshops which addressed wider clinical trial design issues. One group (Smith et al. [26]) aimed to ascertain the opinions of clinical experts about which

Table 2. Summary of Included Studies

Speciality	Condition	Authors
Critical care	Sepsis	Goldstein 2005 [18]
Dentistry	Dental restoration	DeRouen 2002 [27] ^a
Gastroenterology	Crohn disease	Griffiths 2005 [11]
Haematology/bone transplant medicine	GVHD	Pavletic 2006 [13]
Neonatology	Apnoea	Finer 2006 [22] ^{b,c}
	Cardiac instability	Short 2006 [24] ^{b,c}
	Pain	Anand 2005 [41], Anand 2006 [20] ^{b,c}
	Postoperative cardiac dysfunction	Roth 2006 [23] ^{b,c}
	Seizures	Clancy 2006 [21] ^{b,c}
Neurology	Infantile spasms	Lux 2004 [25], Osborne (on behalf of the West Delphi group) 2001 [42]
Psychiatry	Bipolar affective disorder	Carlson 2003 [17]
	Nonepileptic seizures	La France 2006 [19] ^b
Respiratory medicine	Asthma	Smith 1996 [26]
	Cystic fibrosis	Ramsey 1994 [12]
Rheumatology	IIM	Miller 2001 [15], Rider 2002 [43], Rider 2003 [33], Rider (on behalf of IMACS) 2004 [44], Oddis 2005 (describing the work of IMACS) [45]
	SLE	Ruperto 2003 [16], Ruperto 2004 [30], Ruperto (on behalf of PRINTO) 2006 [46]
	Rheumatoid arthritis	Giannini 1997 [14]

^aConcerned the selection of outcomes for use in a specific clinical trial.

^bConcerned the selection of outcomes which could be used in clinical trials of children with clinical features of an underlying condition.

^cCollaboration working on behalf of the Neonatal Drug Development Initiative (NDDI).

doi:10.1371/journal.pmed.0050096.t002

outcomes to measure in clinical trials in children with asthma. One group (DeRouen et al. [27]) ascertained the opinions of experts about which outcome to measure in a specific safety trial of two interventions used in paediatric dental restoration. Our search identified no systematic reviews of studies that had selected outcome measures for use in clinical trials.

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

Methodological Quality of Studies

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

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

sufficient detail to be able to repeat the study (Smith et al. [26]).

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

Methods Used to Select Outcomes

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

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

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

mented on their suitability and content. Rounds five and six consisted of formulation of a draft and, subsequently, a final paper that were altered following comments from the group.

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

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

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

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

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

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

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

People Involved in Selecting Outcomes

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

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

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

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

Techniques Used to Validate Outcomes

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

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

The IMACS group retrospectively assessed the validity,

reliability, and responsiveness of the outcomes they had selected by reviewing available literature on the topic [15].

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

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

Which Outcomes Were Selected by the Groups?

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

Discussion

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

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

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

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

encouraged. The difficulties of undertaking this task, however, should not be underestimated [31].

Robustness and Limitations of the Review

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

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

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

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

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

Core Outcomes

If implemented, the studies we have identified should reduce the impact of inappropriate outcome selection on the

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

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

European Drug Regulation

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

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

The Selection of Outcomes for Use in Children

It is appropriate that some aspects of study design in clinical trials in children differ from equivalent studies performed in adults, and selection of outcomes is one such issue that trialists should consider. In certain situations outcome selection may be similar between the two groups, and some outcomes could be appropriately transposed from adult studies into trials in paediatric populations either in their original state or with slight modification. The danger,

however, of not acknowledging the differences between children and adults with the same disease is that the overall validity of trial results could be compromised. Griffiths et al. [11], when discussing which outcomes to measure in clinical trials of children with Crohn disease, highlight that the importance of linear growth is “unique to pediatric patients.” Another example of an outcome exclusive to children is the assessment of neurodevelopment. Other differences between adults and children that may preclude the use of the same outcomes for both groups include distinct disease pathogenesis, different clinical features and natural history, variations in physiological and psychological outcomes, and contrasting roles within the contexts of families and society in general that may preclude the use of the same outcomes.

Unanswered Questions

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

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

Conclusions

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

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

Supporting Information

Table S1. Search Methods for Identification of Studies

Found at doi:10.1371/journal.pmed.0050096.st001 (114 KB DOC).

Table S2. Characteristics of Excluded Studies

Found at doi:10.1371/journal.pmed.0050096.st002 (243 KB DOC).

Table S3. Characteristics of Included Studies

Found at doi:10.1371/journal.pmed.0050096.st003 (145 KB DOC).

Table S4. Critical Appraisal of the Methodological Quality of Included Studies

Found at doi:10.1371/journal.pmed.0050096.st004 (110 KB DOC).

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

Found at doi:10.1371/journal.pmed.0050096.st005 (156 KB DOC).

Text S1. QUOROM Checklist

Found at doi:10.1371/journal.pmed.0050096.sd001 (41 KB DOC).

Acknowledgments

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

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

References

- Suls J, Rothman A (2004) Evolution of the biopsychosocial model: prospects and challenges for health psychology. *Health Psychol* 23: 119–125.
- Jones B (1987) Impairment, disability and handicap. *Child Care Health Dev* 13: 359.
- Nosadini R, Tonolo G (2004) Relationship between blood glucose control, pathogenesis and progression of diabetic nephropathy. *J Am Soc Nephrol* 15: S1–S5.
- Molenberghs G, Burzykowski T, Alonso A, Buyse M (2004) A perspective on surrogate endpoints in controlled clinical trials. *Stat Methods Med Res* 13: 177–206.
- Tugwell P, Boers M (1993) OMERACT conference on outcome measures in rheumatoid arthritis clinical trials: Introduction. *J Rheum* 528–530.
- Fleming TR, DeMets DL (1996) Surrogate end points in clinical trials: Are we being misled? *Ann Intern Med* 125: 605–613.
- Holloway RG, Dick AW (2002) Clinical trial end points: On the road to nowhere? *Neurology* 58: 679–686.
- Clarke M (2007) Standardising outcomes for clinical trials and systematic reviews. *Trials* 8: 39.
- Duncan PW, Jorgensen HS, Wade DT (2000) Outcome measures in acute stroke trials: A systematic review and some recommendations to improve practice. *Stroke* 31: 1429–1438.
- Giacoa GP, Birenbaum DL, Sachs HC, Mattison DR (2006) The newborn drug development initiative. *Pediatrics* 117: S1–S8.
- Griffiths AM, Otley AR, Hyams J, Quiros AR, Grand RJ, et al. (2005) A review of activity indices and end points for clinical trials in children with Crohn's disease. *Inflamm Bowel Dis* 11: 185–196.
- Ramsay BW, Boat TF (1994) Outcome measures for clinical trials in cystic fibrosis. Summary of a Cystic Fibrosis Foundation consensus conference. *J Pediatr* 124: 177–192.
- Pavletic SZ, Martin P, Lee SJ, Mitchell S, Jacobsohn D, et al. (2006) 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. *Biol Blood Marrow Transplant* 12: 252–266.
- Giannini EH, Ruperto N, Ravelli A, Lovell DJ, Felson DT, et al. (1997) Preliminary definition of improvement in juvenile arthritis. *Arthritis Rheum* 40: 1202–9.
- Miller FW, Rider LG, Chung YL, Cooper R, Danko K, et al. (2001) Proposed preliminary core set measures for disease outcome assessment in adult and juvenile idiopathic inflammatory myopathies. *Rheumatology* 40: 1262–1273.
- Ruperto N, Ravelli A, Murray KJ, Lovell DJ, Andersson-Gare B (2003) Preliminary core sets of measures for disease activity and damage assessment in juvenile systemic lupus erythematosus and juvenile dermatomyositis. *Rheumatology* 42: 1452–1459.
- Carlson GA, Jensen PS, Findling RL, Meyer RE, Calabrese J, et al. (2003) Methodological issues and controversies in clinical trials with child and adolescent patients with bipolar disorder: Report of a consensus conference. *J Child Adolesc Psychopharmacol* 13: 13–27.
- Goldstein B, Giroir B, Randolph A, Members of the International Consensus Conference on Pediatric Sepsis (2005) International pediatric sepsis consensus conference: Definitions for sepsis and organ dysfunction in pediatrics. *Pediatr Crit Care Med* 6: 2–8.
- LaFrance WC Jr., Alper K, Babcock D, Barry JJ, Benbadis S, et al. (2006) Nonepileptic seizures treatment workshop summary. *Epilepsy Behav* 8: 451–461.
- Anand KJ, Aranda JV, Berde CB, Buckman S, Capparelli EV, et al. (2006) Summary proceedings from the neonatal pain-control group. *Pediatrics* 117: S9–S22.
- Clancy RR (2006) Summary proceedings from the neurology group on neonatal seizures. *Pediatrics* 117: S23–S27.
- Finer NN, Higgins R, Kattwinkel J, Martin RJ (2006) Summary proceedings from the apnea-of-prematurity group. *Pediatrics* 117: S47–S51.
- Roth SJ, Adatia I, Pearson GD, Members of the Cardiology Group (2006) Summary proceedings from the cardiology group on postoperative cardiac dysfunction. *Pediatrics* 117: S40–S46.
- Short BL, Van MK, Evans JR (2006) Summary proceedings from the cardiology group on cardiovascular instability in preterm infants. *Pediatrics* 117: S9.
- Lux AL, Osborne JP (2004) A proposal for case definitions and outcome measures in studies of infantile spasms and West syndrome: consensus statement of the West Delphi group. *Epilepsia* 45: 1416–1428.
- Smith MA, Leeder SR, Jalaludin B, Smith WT (1996) The asthma health outcome indicators study. *Aust N Z J Public Health* 20: 69–75.
- DeRouen TA, Leroux BG, Martin MD, Townes BD, Woods JS, et al. (2002) Issues in design and analysis of a randomized clinical trial to assess the safety of dental amalgam restorations in children. *Control Clin Trials* 23: 301–320.
- Dalkey N (1969) The Delphi method: An experimental study of group opinion. Santa Monica (California): Rand.
- van Teijlingen E, Pitchforth E, Bishop C, Russell E (2006) Delphi method and nominal group technique in family planning and reproductive health research. *J Fam Plann Reprod Health Care* 32: 249–252.
- Ruperto N, Ravelli A, Cuttica R, Espada G, Ozen S, et al. (2005) The Pediatric Rheumatology International Trials Organization criteria for the evaluation of response to therapy in juvenile systemic lupus erythematosus: prospective validation of the disease activity core set. *Arthritis Rheum* 52: 2854–2864.
- Lewis A (1992) Group child interviews as a research tool. *Br Educ Res J* 18: 413–421.
- Devane D, Begley CM, Clarke M, Horey D, O'Boyle C (2007) Evaluating maternity care: A core set of outcome measures. *Birth* 34: 164–172.
- Rider LG, Giannini EH, Harris-Love M, Joe G, Isenberg D, et al. (2003) Defining clinical improvement in adult and juvenile myositis. *J Rheum* 30: 603–617.
- Hutton JL, Williamson PR (2000) Bias in meta-analysis due to outcome variable selection within studies. *R Stat Soc Ser C Appl Stat* 49: 359–370.
- Williamson PR, Gamble C, Altman DG, Hutton JL (2005) Outcome selection bias in meta-analysis. *Stat Methods Med Res* 14: 515–524.
- Chan AW, Altman DG (2005) Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *BMJ* 330: 753.
- Connors C, Epstein J, March J, Angold A, Wells K (2001) Multimodal treatment of ADHD in the MTA: An alternative outcome analysis. *J Am Acad Child Adolesc Psychiatry* 40: 159–167.
- Juniper EF, Guyatt GH, Feeny DH, Griffith LE, Ferrie PJ (1997) Minimum skills required by children to complete health-related quality of life instruments for asthma: Comparison of measurement properties. *Eur Respir J* 10: 2285–2294.
- Raat H, Botterweck AM, Landgraf JM, Hoogveen WC, Essink-Bot ML (2005) Reliability and validity of the short form of the child health questionnaire for parents (CHQ-PF28) in large random school based and general population samples. *J Epidemiol Community Health* 59: 75–82.
- Deal L, Gold BD, Gremse DA, Winter HS, Peters SB, Fraga PD, Mack ME, Gaylord SM, Tolia V, Fitzgerald JF (2005) Age-specific questionnaires distinguish GERD symptom frequency and severity in infants and young children: development and initial validation. *J Pediatr Gastroenterol Nutr* 41: 178–185.
- Anand KJS, Aranda JV, Berde CB, Buckman S, Capparelli EV (2005) Analgesia and anesthesia for neonates: study design and ethical issues. *Clin Ther* 27: 814–843.
- Osborne JP, Lux A (2001) Towards an international consensus on definitions and standardised outcome measures for therapeutic trials (and epidemiological studies) in West syndrome. *Brain Dev* 23: 677–682.
- Rider LG (2002) Outcome assessment in the adult and juvenile idiopathic inflammatory myopathies. *Rheum Dis Clin North Am* 28: 935–977.
- Rider LG, Giannini EH, Brunner HI, Ruperto N, James-Newton L, et al. (2004) International consensus on preliminary definitions of improvement in adult and juvenile myositis. *Arthritis Rheum* 50: 2281–2290.
- Oddis CV (2005) Outcomes and disease activity measures for assessing treatments in the idiopathic inflammatory myopathies. *Curr Rheum Rep* 7: 87–93.
- Ruperto N, Ravelli A, Oliveira S, Alessio M, Mihaylova D (2006) The Pediatric Rheumatology International Trials Organization/American College of Rheumatology provisional criteria for the evaluation of response to therapy in juvenile systemic lupus erythematosus: prospective validation of the definition of improvement. *Arthritis Rheum* 55: 355–363.

Editors' Summary

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

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

What Did the Researchers Do and Find? The researchers used a predefined search strategy to identify all the studies published since 1950 that examined the selection of outcomes in clinical trials in children. They also asked experts in pediatric clinical research for details of relevant studies. Only 25 studies, which covered several pediatric specialties and were published by 13 collaborative groups, met the strict eligibility criteria laid down by the researchers for their systematic review. Several approaches previously used to choose outcomes in clinical trials in adults were used in these studies to select outcomes. Two

groups used the “Delphi” technique, in which opinions are sought from individuals, collated, and fed back to the individuals to generate discussion and a final, consensus agreement. One group used the “nominal group technique,” which involves the use of structured face-to-face discussions to develop a solution to a problem followed by a vote. Another group used both methods. The remaining groups (except one that used a questionnaire) used semistructured discussion meetings or workshops to decide on outcomes. Although most of the groups included clinical experts, people doing research on the specific clinical condition under investigation, and industry representatives, only three groups asked parents about which outcomes should be included in the trials, and none asked children directly.

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

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

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