PEDIATRICS®
OFFICIAL JOURNAL OF THE AMERICAN ACADEMY OF PEDIATRICS

Standard 4: Determining Adequate Sample Sizes

Ingeborg van der Tweel, Lisa Askie, Ben Vandermeer, Susan Ellenberg, Ricardo M. Fernandes, Haroon Saloojee, Dirk Bassler, Douglas G. Altman, Martin Offringa and Johanna H. van der Lee

Pediatrics 2012;129;S138
DOI: 10.1542/peds.2012-0055G

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://pediatrics.aappublications.org/content/129/Supplement_3/S138.full.html

PEDIATRICS is the official journal of the American Academy of Pediatrics. A monthly publication, it has been published continuously since 1948. PEDIATRICS is owned, published, and trademarked by the American Academy of Pediatrics, 141 Northwest Point Boulevard, Elk Grove Village, Illinois, 60007. Copyright © 2012 by the American Academy of Pediatrics. All rights reserved. Print ISSN: 0031-4005. Online ISSN: 1098-4275.

American Academy of Pediatrics
DEDICATED TO THE HEALTH OF ALL CHILDREN™

Downloaded from pediatrics.aappublications.org at University of Manitoba Libraries on April 9, 2014

Standard 4: Determining Adequate Sample Sizes

DILEMMA

There are many challenges to be faced when conducting randomized controlled trials (RCTs) in pediatric research. One important challenge is the determination of an appropriate sample size. Recruiting more children than necessary risks unnecessary overexposure of children to an inferior treatment, whereas underestimating the sample required will lead to inconclusive or unreliable results. Both options pose important ethical dilemmas for the pediatric researcher.

Reviews have concluded that sample size calculations are frequently based on inaccurate assumptions regarding the key pieces of information needed.[1,2] For example, a recent "failed" pediatric RCT highlighted the commonly encountered problem of underestimating the SD of a continuous primary outcome variable.[3] This leads to underestimation of the necessary sample size, inadequate statistical power, and, consequently, an unanswered study question. Similarly, an incorrect estimation of the frequency with which a dichotomous outcome, or event, occurs in the control group of a trial (known as the control or baseline event rate) will also lead to incorrect sample size estimation. This standards article uses a series of scenarios to assist pediatric researchers in not only determining an adequate trial sample size but also how to proceed when this sample size may be difficult to achieve. Recommendations for practice are summarized in Table 1.

GUIDANCE

Methods to calculate adequate sample sizes are described for superiority, noninferiority, and cluster-randomized trials with various types of primary outcome variables, including continuous, dichotomous, or time-to-event data.[4–10] In this article, we focus primarily on superiority trial designs with continuous or dichotomous outcomes. Information on noninferiority or equivalence designs, as well as on cluster-randomized designs, can be found in the literature.[9,10] RCTs generally have many outcomes of interest, but the number of patients required (and the main statistical analysis) should be based on the primary outcome.

A sample size calculation for a standard (2-sided, superiority) RCT is based on 3 values. First, one must specify the target difference between the outcomes for those children who received the new treatment compared with those who did not, based on what is judged clinically relevant or meaningful. Next, one must specify the level of risk one is willing to take that, by chance, the trial will erroneously conclude a clinically relevant difference in the primary outcome. This is known as the type I error rate or α (the probability of a false-positive conclusion). Finally, one must specify the probability that the trial will correctly detect a difference in the primary outcome between the control and

AUTHORS: Ingeborg van der Tweel, PhD,[a] Lisa Askie, PhD, MPH, BN, NICC, RM, RN,[b] Ben Vandermeer, MSc,[c] Susan Ellenberg, PhD,[d] Ricardo M. Fernandes, MD,[e] Haroon Saloojee, MD, MBBCh, FCPaed,[f] Dirk Bassler, MD, MSC,[g] Douglas G. Altman, DSc,[h] Martin Offringa, MD, PhD,[i] and Johanna H. van der Lee, MD, PhD,[j] for the StaR Child Health Group

[a]Biostatistics, Julius Centre for Health Sciences and Primary Care, University Medical Centre, Utrecht, Netherlands; [b]NHMRC Clinical Trials Centre, University of Sydney, Sydney, Australia; [c]Department of Pediatrics, Alberta Research Centre for Health Evidence, University of Alberta, Edmonton, Alberta, Canada; [d]Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania; [e]Department of Pediatrics, Santa Maria Hospital, Laboratory of Clinical Pharmacology and Therapeutics, Faculty of Medicine, Instituto de Medicina Molecula, Lisboa, Portugal; [f]Division of Community Pediatrics, Department of Pediatrics and Child Health, University of the Witwatersrand, Johannesburg, South Africa; [g]Centre for Pediatric Clinical Studies and Department and Department of Neonatology, University Children's Hospital, Tuebingen, Germany; [h]Centre for Statistics in Medicine, University of Oxford, Oxford, United Kingdom; [i]Child Health Evaluative Sciences, Research Institute, The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada; and [j]Department of Pediatric Clinical Epidemiology, Emma Children's Hospital, Academic Medical Centre, Amsterdam, Netherlands

KEY WORDS

Bayesian methods, nuisance estimates, pediatric, sample size, StaR Child Health

ABBREVIATIONS

CINV—chemotherapy-induced nausea and vomiting
RCT—randomized controlled trial
RSV—respiratory syncytial virus

Drs van der Tweel, Askie, and Vandermeer wrote the first draft of the article; Drs Ellenberg, Fernandes, Saloojee, Bassler, Altman, Offringai, and van der Lee contributed to the writing of the article; Drs van der Tweel, Askie, Vandermeer, Ellenberg, Fernandes, Saloojee, Bassler, Altman, and van der Lee participated in regular meetings and conference calls, identified the issues, and drafted the manuscript; Drs van der Tweel, Askie, Vandermeer, Ellenberg, Fernandes, Saloojee, Bassler, Altman, Offringa, and van der Lee participated in identifying the evidence base for StaR Child Health standards; and Drs van der Tweel, Askie, Vandermeer, Ellenberg, Fernandes, Saloojee, Bassler, Altman, Offringa, and van der Lee agree with the final version.

This is the fourth in a series of standard articles resulting from an ongoing process in which a group of invited experts called a Standard Development Group from StaR Child Health assembles and exchanges information about methods for pediatric trial design, conduct, and reporting. More detailed information about this topic can be found in the introductory article of this supplement or at the StaR Child Health Web site (www.starchildhealth.org).

(Continued on last page)

Downloaded from pediatrics.aappublications.org at University of Manitoba Libraries on April 9, 2014

TABLE 1 Recommendations for Practice

- Sample size calculation is a matter of Good Clinical Practice in the design phase of an RCT
- When previous information is available from similar pediatric populations, sample size calculation can be performed using standard methods or software
- When the only previous information is from adult or other pediatric populations or when no previous information is available, alternative methods should be considered
- Parameter values used for the sample size calculation (eg, type I error, power, SD, control event rate, expected treatment difference) should be clearly mentioned in the final trial report
- Good follow-up procedures are essential so children are not lost to follow up
- Always consult a statistician or methodologist about the sample size calculation during the planning phase of a trial

the new intervention if the true difference is the size of the target difference. This is known as the statistical power (the probability of a true positive conclusion). Conventional values for α and power are .05 (or 5%) and 80% to 90%, respectively. Requiring a smaller probability of a type I error, larger power, or smaller clinically relevant difference will increase the necessary sample size. A smaller value for α may be appropriate to reduce the chance of a false-positive conclusion when an effective standard therapy is tested against a new, competing treatment. A higher power could be considered if, in comparing treatments to prevent a common disease, one would not miss a safe, inexpensive, possibly effective treatment.[11]

To make a sample size calculation, so-called nuisance parameters also need to be estimated. In statistics, a nuisance parameter is any parameter which is not of immediate interest but that must be accounted for in the analysis of those parameters which are of interest. For example, for continuous outcomes, such as blood pressure or duration of mechanical ventilation, an estimate of the SD is needed. For dichotomous outcomes, such as dead/alive, an accurate estimate of the control or baseline event rate is needed. To estimate the values for these nuisance parameters, information is needed from comparable previous studies. Ideally, such studies would have included similar populations of children; however, this is often not the case.

The following scenarios describe situations that may be encountered when determining sample size for a pediatric clinical trial and possible approaches. The overall approach is summarized in Fig 1.

SCENARIO 1

You already have the required information from other similar pediatric populations. How do you calculate the required sample size for your new trial?

Possible Approaches

If information is available from previous studies for a reliable estimate of the nuisance parameters, then a standard sample size calculation can be made based on whatever clinically meaningful difference is considered important to detect. This can be done by using methods previously described[4–10] or using readily available software.

Example 1

There have been 2 previous small trials of corticosteroids to treat respiratory syncytial virus (RSV) lower respiratory tract infections in young children, but more evidence is needed. These 2 trials found similar reductions in the primary outcome (ie, hours of mechanical ventilation required). Thus, a possible approach would be to calculate the SD estimate based on the average SD from these 2 trials. In this case, the average SD from the 2 trials was 84 hours. A reduction of 36 hours' ventilation with corticosteroid treatment compared with placebo is considered clinically relevant. Assuming a 2-sided type I error of .05 (5%) and power of 80%, by entering these numbers into a standard sample size software package, it can be determined that the minimum required sample size for a new trial would be 86 children in each arm.

If information is available from several comparable (but often small) RCTs or subgroups within RCTs, these can be combined in a more systematic and quantitative way, known as meta-analysis, to obtain an estimate of a nuisance parameter.[12–14] Anyone planning to conduct a new trial should first undertake a thorough review of the available literature (known as a systematic review).[12] The information from the identified studies can be entered into readily available meta-analysis software to obtain estimates of various nuisance parameters.[15] These parameters are obtained by looking at the event rate for the combined control groups for dichotomous outcomes or the SD for the combined control groups for continuous outcomes. If the evidence is limited, such as in Example 1 with the 2 small trials, one might be more conservative and take a higher value than just the average. It is further recommend to examine how sensitive the sample size is to variation in the nuisance parameter.[16]

Example 2

For decades, methylxanthines (eg, caffeine, theophylline, aminophylline) have been commonly given to preterm infants to reduce their rates of apnea. In 2001, a systematic review indicated that this treatment had only been assessed in 5 small trials with a total of 192 randomized infants.[17] Using this information, a group of pediatric trialists estimated that the number of infants needed to be randomized to provide definitive evidence of both long- and

Downloaded from pediatrics.aappublications.org at University of Manitoba Libraries on April 9, 2014

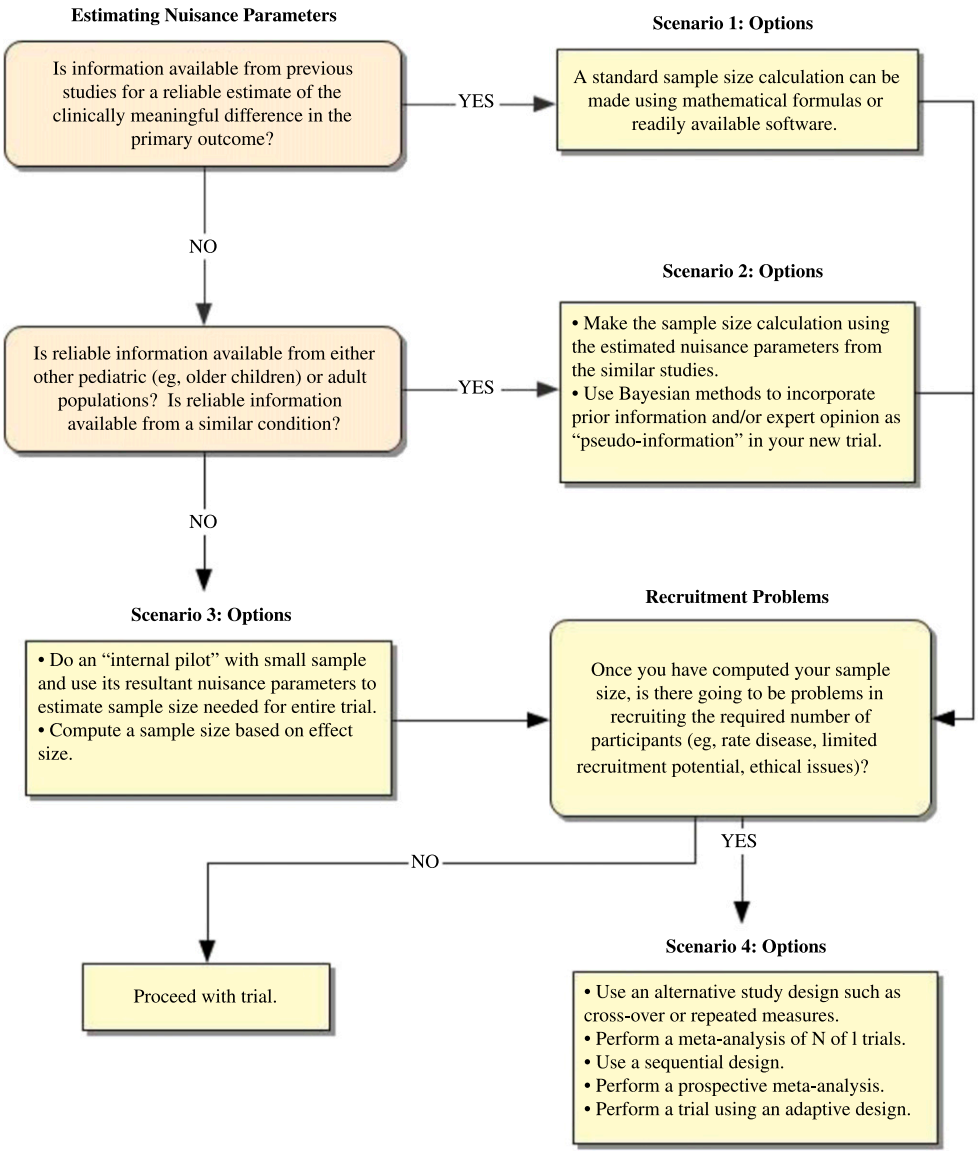

FIGURE 1
Determining sample size.

short-term benefits (and potential harms) would be ~2000. This finding was based on an estimate of the baseline event rate of death or neurodevelopment disability (the primary outcome) of 20% and having 80% power to detect a 25% relative reduction in the risk of this outcome. A multicenter international trial subsequently randomized 2006 infants to treatment and demonstrated conclusively that caffeine compared with placebo significantly reduced the risk of death or disability in preterm infants.[18]

SCENARIO 2

The only previous information you have is from either adult populations or pediatric populations that differ from your group of interest. How do you calculate the required sample size for your new trial?

Possible Approaches

In this scenario, the most common available sources of nuisance estimates are studies of similar design evaluating similar treatments, even if the population studied differs from the type of children you are interested in. The 2 examples that follow here show how information obtained from either other pediatric populations (Example 3) or adults (Example 4) can inform a sample size calculation for a new trial.

Example 3

Following on from Example 1, you wish to investigate, in a placebo-controlled trial, whether dexamethasone is effective in reducing the duration of mechanical

Downloaded from pediatrics.aappublications.org at University of Manitoba Libraries on April 9, 2014

ventilation in children aged <24 months with severe RSV lower respiratory tract infections. The best available data are from a trial published 6 years ago that used a different type of corticosteroid and included children admitted to the hospital with RSV bronchiolitis (both ventilated and nonventilated children). Based on data of a subgroup of ventilated children from this trial, the SD was estimated as 98 hours. By using these estimates, if you designed a new trial with 80% power and a 2-sided α of .05, you would need to recruit, at minimum, 117 children to each arm of the trial to detect a difference of 36 ventilation hours.

Example 4

Ginger therapy has been effective in reducing chemotherapy-induced nausea and vomiting (CINV) in adults. However, no data were available regarding the antiemetic efficacy of ginger in pediatric patients receiving chemotherapy. A new trial is planned starting with the adult information, which showed a CINV rate ranging from 30% to 90%.[19] If a relative reduction in the rate of CINV of 20% is considered clinically relevant (this is another assumption, which may vary because the participants are children rather than adults), and a type I error of .05 and power of 80% is assumed, the sample size calculation will depend greatly on the control event rate. If a control event rate of 60% is assumed (approximately the middle of the 30%–90% range found in adults), then a new trial would require at least 267 patients in each arm. However, if the maximal control event rate found in adults (90%) is assumed, the new pediatric trial would require only 71 patients per arm.

Another way of addressing the issue of using information from adult or other pediatric populations is to take a Bayesian approach.[20–22] This approach incorporates previous information in the assessment of data from the new trial. Previous information may be sourced from previous trial data and/or expert opinion and will be combined with the new trial data to yield effect estimates. With this approach, a reduction of the number of trial participants to include can often be expected.

Example 5

Bayesian methods were used to obtain a sample size in a trial comparing intravenous immunoglobulin and plasmapheresis in treating Guillain-Barré syndrome in children.[21] The authors showed that using standard methods to obtain a sample size for their noninferiority design would have required a sample size beyond what was reasonably feasible. By using previous information based on empirical data from adult RCTs combined with expert opinion, they were able to greatly reduce the required sample size.

However, if the previous information is incorrectly specified or trial data are not consistent with previous information, an over- or underpowered study may still result, as is the case in all sample size calculations. This problem can be exacerbated when the previous information comes from studies conducted in adults. Caution must be exercised when translating previous information from adults or specific pediatric populations to (other types of) children, as there may be differences in disease definitions/criteria and outcomes, as well as differences in biology, physiology, pharmacology, and measurement. Importantly, children may have larger variations in some outcomes than adults, especially when a wide age range is included in the study. The issues regarding the use of appropriate lung function reference equations in trials of asthma and other respiratory conditions are 1 such example.[23] Better solutions are needed to help combat this problem. If one is not confident about the validity of estimates from these existing studies, it may be better to proceed to scenario 3, rather than use inaccurate estimates.

SCENARIO 3

The previous information needed to calculate a pediatric trial sample size is either not available from relevant pediatric or adult populations or any information that is available is considered too unreliable for use. How do you calculate the required sample size for your new trial?

Possible Approaches

An internal pilot is 1 possible solution. First, the nuisance parameter (eg, SD) is estimated for the primary outcome from a previous adult trial. With this estimate, a preliminary sample size calculation can be made. A pilot phase for a trial can then be commenced in children, and after a prespecified number of patients are included, the nuisance parameter can be re-estimated on the basis of available information. This information can be used to reassess the final sample size requirements. The first part of the data are considered an internal pilot (ie, the participants enrolled in this phase of the trial will be part of the final sample size). However, because there has been an interim "look" at the data, there is a risk of inflating the trial's type I error. If the interim look is used only to estimate the nuisance parameter from the control group data and the treatment difference is not estimated, then this inflation will be negligible.[24–28]

Example 6

If, in Example 3, you considered the previous trial information to be too unreliable, you could undertake an internal pilot study of 30 children in each arm. Suppose, from the blinded pilot data, the common SD was 108 hours

Downloaded from pediatrics.aappublications.org at University of Manitoba Libraries on April 9, 2014

(not 98 hours as had been estimated in the previous trial). Based on these pilot data, your re-estimated sample size is at least 142 children in each arm of the trial to detect a difference of 36 hours in ventilation duration. Thus, you would need to recruit another 112 children in each arm in addition to those already included in the pilot study.

Another possibility if no previous estimate for the SD can be derived is to use the effect size to calculate the necessary sample size. The effect size is the standardized treatment difference. For a continuous outcome, this is the ratio of the treatment difference to the SD. Cohen proposed values for small, moderate, and large effect sizes for various types of outcomes.[7] However, as effect sizes are abstract concepts, it is sometimes difficult to translate them into clinically interpretable quantities.

Example 7

Suppose no information was available to estimate the SD in the trial described in Example 3. We could design a trial with 80% power and a 2-sided α of .05 to detect a standardized difference of .5, which, according to Cohen, would be a "moderate" effect.[7] This may be a reasonable assumption (rather than assuming a very small or large effect size) given no available previous information is available. With these assumptions, at least 63 patients would need to be recruited in each arm. Note that the effect size calculated in Example 3 would have been 36 of 98 (ie, 0.37), which could be considered small to moderate.

SCENARIO 4

The number of children required to obtain reliable trial results is large, but the number of children potentially available for inclusion in the trial is limited (eg, a rare disease). What are the available options?

Possible Approaches

The European Medicines Agency guideline on clinical trials in small populations[29] states that "… no methods exist that are relevant to small studies that are not also applicable to large studies. However, it may be that in conditions with small and very small populations, less conventional and/or less commonly seen methodological approaches may be acceptable if they help to improve the interpretability of the study results." Some suggested approaches[29,30] for addressing the issue of limited availability of children for inclusion in a trial are given here.

First, one could switch to another (type of) outcome, but only if this also represents a clinically relevant outcome. Another approach could be to extend the follow-up time so more events will accrue.

Crossover Design

A simple method for reducing the sample size needed is to use a different design. A crossover study exposes all participants to both the new intervention and the control treatment. This method is only suitable in specific circumstances; namely, when evaluating interventions with a temporary effect in the treatment of stable, chronic conditions and when the outcomes are short-term and reversible (eg, headache relief). It is also important to allow enough time to pass between treatment exposures to ensure that there is no carryover effect from 1 treatment to the other.[31]

Repeated Measures

The sample size needed can also be reduced by undertaking repeated measures of the primary outcome in each trial patient. Multiple observations per patient will reduce the number of patients required.[32] Again, however, this trial design can only be used if an outcome can be measured several times per patient (eg, blood pressure, quality of life). Furthermore, repeated measures require relatively sophisticated statistical design and analysis.

Meta-analysis of N-of-1 Trials

Another potential approach for addressing small available sample sizes in rare diseases is to meta-analyze "N-of-1 trials." N-of-1 trials include only 1 participant per trial who acts as his or her own control, using a crossover approach. Typically, the participant has repeated measures of the same outcome. Various modeling methods have been developed to combine (meta-analyze) the results of such trials.[33] This approach could be useful in designing pediatric pilot studies, when funding and/or participants are difficult to obtain. However, this design will only be feasible when dealing with short-term temporary outcomes (eg, pain) and no carryover effect.[34]

Sequential Designs

Most RCTs estimate the sample size in the design phase of the trial and subsequently include, randomize, and follow up their patients until the primary outcome values are obtained for all participants. Trials using sequential designs provide for regular analyses of interim data with the possibility of terminating the trial before full accrual has been achieved if the data reach a predetermined threshold for making a definitive judgment about the results. It is widely recognized that an "adjustment" of the type I error for each of the interim analyses is required to avoid increasing the probability of a false-positive conclusion if the trial were to be stopped early on the basis of an interim analysis. Various ways of adjustment have been described.[33–36]

Downloaded from pediatrics.aappublications.org at University of Manitoba Libraries on April 9, 2014

Although sequential designs may on occasion lead to a smaller sample size when the interim data demonstrate definitively that 1 treatment is better than another, if there are serious safety concerns or further study is unlikely to demonstrate any difference, such designs cannot be relied on to reduce the sample size.

Prospective Meta-analysis

Traditionally, when a large sample size is required, a multicenter trial is often needed to ensure sufficient numbers of children are recruited in a reasonable period of time (see Example 2). However, conducting a large, multicenter trial is not always possible due to funding, feasibility, and timeline constraints. A possible solution is to prospectively plan a series of individual trials that all follow identical (or very similar) protocols but exist independently of 1 another. The trialists form a collaborative group and prospectively agree to collect core, key data items in a common format and also agree to share these data on trial completion for inclusion in a combined meta-analysis. This is known as a prospective meta-analysis.[37] By agreeing to harmonize the data items, definitions, coding, and standards before the data are collected, the prospective meta-analysis trialists guarantee that the amount of data available for later synthesis is maximized.

Example 8

Four trials are currently underway to assess whether interventions commenced very early in infancy reduce childhood obesity. Each of the 4 trials (with between 400 and 800 children enrolled in each) has sufficient statistical power to detect meaningful differences in BMI *z* scores at 2 years of age. However, it is only with the combined sample size from the 4 trials of 1800 children that the outcome of most public health significance (ie, a reduction in obesity prevalence from 20% to 15%) will be able to be reliably detected. These trialists have thus formed the Early Prevention of Obesity in Children Collaboration[38] to undertake a prospective meta-analysis to achieve the required sample size and answer the primary research question.

Adaptive Designs

Adaptive designs are flexible designs that permit modifications during the course of the trial. The flexibility covers a wide range of possible adaptations, including changes in sample size. For example, the internal pilot design discussed earlier can be viewed as an adaptive design, as can sequential designs. The price paid for this flexibility is that the implementation of such designs is often complex and, in some cases, can reveal interim trends that optimally should be kept confidential during the study.[39,40] Adaptive designs can also lead to larger sample sizes than originally estimated. These factors may complicate assessments of the overall trial cost and should thus be taken into consideration when deciding whether to use these methods during the trial's planning phase.

RESEARCH AGENDA

This Standard Development Group has identified a number of areas for future research related to determining sample sizes in pediatric research. An evaluation of the use of information from adult studies for pediatric trial designs would clarify situations or conditions in which this type of extrapolation could lead to unacceptable errors in sample size calculation. Although the use of Bayesian methods and adaptive designs may reduce the overall sample size, such designs can also increase the required sample size in some cases, possibly presenting complications to researchers and funders. Further research should investigate when these methods may be most appropriate. The implications of a prospective meta-analysis have yet to be fully appreciated. Specifically, it is not clear how results may differ from a multicenter trial on the same topic. Although the prospective nature of the collaboration overcomes many potential limitations of multiple separate studies, exploration of the benefits and limitations of this type of research endeavor will be important.

CONCLUSIONS

This standards article provides an introduction to a variety of methods to overcome common challenges faced when attempting to derive sample size estimates in pediatric research. The desire to reduce exposure of trial participants to any potential harm must be carefully balanced with the need to obtain an adequate sample size to ensure the validity of trial results. Fortunately, there are well-described methods to determine sample size that capitalize on previous data, whether from a similar or different population, or on a systematic review. To supplement data derived from different populations, internal pilot studies or a Bayesian approach can be used to integrate new data and/or expert opinion into the sample size calculation. Crossover studies, repeated measures, and the use of effect sizes can overcome a lack of previous data or a very small potential subject pool. The use of sequential designs may allow early termination of trials when results are definitive before full accrual is achieved but cannot be relied on as a way to conduct a smaller trial. Such designs should weigh limiting the exposure of trial participants to potentially harmful treatments

Downloaded from pediatrics.aappublications.org at University of Manitoba Libraries on April 9, 2014

with ensuring that the final results are sufficient for reliable benefit-to-risk assessments. Finally, the use of a prospective meta-analysis allows researchers to collectively contribute their data to achieve the necessary power in the absence of resources or possibilities for a multicenter trial.

Ultimately, appropriate time and attention must be dedicated to the determination of sample size in the planning of any trial, and pediatric trials are no exception. Consultation with a methodologist or statistician in the design phase can minimize the number of children exposed to the potential risks of study participation while ensuring the validity of the research findings.

REFERENCES

1. Charles P, Giraudeau B, Dechartres A, Baron G, Ravaud P. Reporting of sample size calculation in randomised controlled trials: review. *BMJ*. 2009;338:b1732
2. Vickers AJ. Underpowering in randomized trials reporting a sample size calculation. *J Clin Epidemiol*. 2003;56(8):717–720
3. van der Lee JH, Tanck MW, Wesseling J, Offringa M. Pitfalls in the design and analysis of paediatric clinical trials: a case of a 'failed' multi-centre study, and potential solutions. *Acta Paediatr*. 2009;98(2):385–391
4. Lachin JM. Introduction to sample size determination and power analysis for clinical trials. *Control Clin Trials*. 1981;2(2):93–113
5. Freedman LS. Tables of the number of patients required in clinical trials using the logrank test. *Stat Med*. 1982;1(2):121–129
6. Donner A. Approaches to sample size estimation in the design of clinical trials—a review. *Stat Med*. 1984;3(3):199–214
7. Cohen J. A power primer. *Psychol Bull*. 1992;112(1):155–159
8. Julious SA. Sample sizes for clinical trials with normal data. *Stat Med*. 2004;23(12):1921–1986
9. Dann RS, Koch GG. Methods for one-sided testing of the difference between proportions and sample size considerations related to non-inferiority clinical trials. *Pharm Stat*. 2008;7(2):130–141
10. Friedman LM, Furberg CD, DeMets DL. *Fundamentals of Clinical Trials*. 4th ed. New York, NY: Springer; 2010
11. Piantadosi S. *Clinical Trials. A Methodological Perspective*. New York, NY: John Wiley & Sons; 1997
12. Clarke M, Hopewell S, Chalmers I. Clinical trials should begin and end with systematic reviews of relevant evidence: 12 years and waiting. *Lancet*. 2010;376(9734):20–21
13. Sutton AJ, Cooper NJ, Jones DR, Lambert PC, Thompson JR, Abrams KR. Evidence-based sample size calculations based upon updated meta-analysis. *Stat Med*. 2007;26(12):2479–2500
14. Whitehead A. *Meta-analysis of Controlled Clinical Trials*. Chichester, West Sussex, United Kingdom: Wiley; 2002
15. Cochrane Collaboration. Review Manager (RevMan). Available at: www.ims.cochrane.org/revman/download. Accessed April 28, 2012
16. Julious SA, Owen RJ. Sample size calculations for clinical studies allowing for uncertainty about the variance. *Pharm Stat*. 2006;5(1):29–37
17. Henderson-Smart DJ, Steer P. Methylxanthine treatment for apnea in preterm infants. *Cochrane Database Syst Rev*. 2001;(3):CD000140
18. Schmidt B, Roberts RS, Davis P, et al; Caffeine for Apnea of Prematurity Trial Group. Long-term effects of caffeine therapy for apnea of prematurity. *N Engl J Med*. 2007;357(19):1893–1902
19. Pillai AK, Sharma KK, Gupta YK, Bakhshi S. Anti-emetic effect of ginger powder versus placebo as an add-on therapy in children and young adults receiving high emetogenic chemotherapy. *Pediatr Blood Cancer*. 2011;56(2):234–238
20. Whitehead J, Valdés-Márquez E, Johnson P, Graham G. Bayesian sample size for exploratory clinical trials incorporating historical data. *Stat Med*. 2008;27(13):2307–2327
21. Goodman SN, Sladky JT. A Bayesian approach to randomized controlled trials in children utilizing information from adults: the case of Guillain-Barré syndrome. *Clin Trials*. 2005;2(4):305–310, discussion 364–378
22. Schoenfeld DA, Hui Zheng , Finkelstein DM. Bayesian design using adult data to augment pediatric trials. *Clin Trials*. 2009;6(4):297–304
23. Stanojevic S, Wade A, Stocks J. Reference values for lung function: past, present and future. *Eur Respir J*. 2010;36(1):12–19
24. Wittes J, Brittain E. The role of internal pilot studies in increasing the efficiency of clinical trials. *Stat Med*. 1990;9(1–2):65–71; discussion 71–72
25. Freidlin B, Korn EL. Release of data from an ongoing randomized clinical trial for sample size adjustment or planning. *Stat Med*. 2007;26(22):4074–4082
26. Kieser M, Friede T. Re-calculating the sample size in internal pilot study designs with control of the type I error rate. *Stat Med*. 2000;19(7):901–911
27. Proschan MA, Liu Q, Hunsberger S. Practical midcourse sample size modification in clinical trials. *Control Clin Trials*. 2003;24(1):4–15
28. Gould AL, Shih WJ. Sample size re-estimation without unblinding for normally distributed outcomes with unknown variance. *Comm Statist Theory Methods*. 1992;21(10):2833–2853
29. Committee for Medicinal Products for Human Use. *Guideline on Clinical Trials in Small Populations*. London, United Kingdom: European Medicines Agency; 2006
30. Evans, Jr, CH and Ildstad ST, eds. Committee on Strategies for Small-Number-Participant. *Small Clinical Trials: Issues and Challenges*. Washington, DC: National Academy Press; 2001
31. Senn S. *Cross-over Trials in Clinical Research*. 2nd ed. Chichester, West Sussex, United Kingdom: Wiley; 2002
32. Winkens B, Schouten HJ, van Breukelen GJ, Berger MP. Optimal number of repeated measures and group sizes in clinical trials with linearly divergent treatment effects. *Contemp Clin Trials*. 2006;27(1):57–69
33. Haybittle JL. Repeated assessment of results in clinical trials of cancer treatment. *Br J Radiol*. 1971;44(526):793–797
34. Peto R, Pike MC, Armitage P, et al. Design and analysis of randomized clinical trials requiring prolonged observation of each patient. I. Introduction and design. *Br J Cancer*. 1976;34(6):585–612

Downloaded from pediatrics.aappublications.org at University of Manitoba Libraries on April 9, 2014

35. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics*. 1979;35(3):549–556
36. DeMets DL, Lan KK. Interim analysis: the alpha spending function approach. *Stat Med*. 1994;13(13–14):1341–1352; discussion 1353–1356
37. Ghersi D, Berlin J, Askie L. Chapter 19: Prospective meta-analysis. In: Higgins J, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011)*. The Cochrane Collaboration; 2011. Available at: www.cochrane-handbook.org. Accessed April 28, 2012
38. Askie LM, Baur LA, Campbell K, et al; EPOCH Collaboration Group. The Early Prevention of Obesity in CHildren (EPOCH) Collaboration—an individual patient data prospective meta-analysis. *BMC Public Health*. 2010;10:728
39. Bauer P, Köhne K. Evaluation of experiments with adaptive interim analyses. *Biometrics*. 1994;50(4):1029–1041
40. Bauer P, Brannath W. The advantages and disadvantages of adaptive designs for clinical trials. *Drug Discov Today*. 2004;9(8):351–357

(Continued from first page)

www.pediatrics.org/cgi/doi/10.1542/peds.2012-0055G

doi:10.1542/peds.2012-0055G

Accepted for publication Mar 23, 2012

Address correspondence to Martin Offringa, MD, PhD, Senior Scientist and Program Head, Child Health Evaluative Sciences, Research Institute, The Hospital for Sick Children, 555 University Ave, Toronto, Ontario, Canada M5G 1X8. E-mail: martin.offringa@sickkids.ca

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).

Copyright © 2012 by the American Academy of Pediatrics

FINANCIAL DISCLOSURE: *The authors have indicated they have no financial relationships relevant to this article to disclose.*

Downloaded from pediatrics.aappublications.org at University of Manitoba Libraries on April 9, 2014

Standard 4: Determining Adequate Sample Sizes

Ingeborg van der Tweel, Lisa Askie, Ben Vandermeer, Susan Ellenberg, Ricardo M. Fernandes, Haroon Saloojee, Dirk Bassler, Douglas G. Altman, Martin Offringa and Johanna H. van der Lee

Pediatrics 2012;129;S138
DOI: 10.1542/peds.2012-0055G

Updated Information & Services	including high resolution figures, can be found at: http://pediatrics.aappublications.org/content/129/Supplement_3/S138.full.html
References	This article cites 29 articles, 4 of which can be accessed free at: http://pediatrics.aappublications.org/content/129/Supplement_3/S138.full.html#ref-list-1
Citations	This article has been cited by 2 HighWire-hosted articles: http://pediatrics.aappublications.org/content/129/Supplement_3/S138.full.html#related-urls
Subspecialty Collections	This article, along with others on similar topics, appears in the following collection(s): **Medical Education** http://pediatrics.aappublications.org/cgi/collection/medical_education_sub
Permissions & Licensing	Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at: http://pediatrics.aappublications.org/site/misc/Permissions.xhtml
Reprints	Information about ordering reprints can be found online: http://pediatrics.aappublications.org/site/misc/reprints.xhtml

PEDIATRICS is the official journal of the American Academy of Pediatrics. A monthly publication, it has been published continuously since 1948. PEDIATRICS is owned, published, and trademarked by the American Academy of Pediatrics, 141 Northwest Point Boulevard, Elk Grove Village, Illinois, 60007. Copyright © 2012 by the American Academy of Pediatrics. All rights reserved. Print ISSN: 0031-4005. Online ISSN: 1098-4275.

American Academy of Pediatrics
DEDICATED TO THE HEALTH OF ALL CHILDREN™

Downloaded from pediatrics.aappublications.org at University of Manitoba Libraries on April 9, 2014