

# Abstracts of Randomized Controlled Trials Presented at the Society for Pediatric Research Meeting

## *An Example of Publication Bias*

Terry P. Klassen, MD; Natasha Wiebe, MMath; Kelly Russell, BSc; Kelly Stevens, BSc; Lisa Hartling, MSc; William R. Craig, MD; David Moher, MSc

**Background:** Publication bias toward studies that favor new therapies has been known to occur for the past 40 years, yet its implications are not well studied in child health. The increased interest in meta-analyses has highlighted the need to identify the totality of evidence when addressing treatment questions.

**Objectives:** To measure the percentage of randomized controlled trials (RCTs) presented at a major pediatric scientific meeting that were subsequently published as full-length articles, to investigate factors associated with publication, and to describe the variables that change from abstract to manuscript form.

**Design:** The scientific proceedings from the Society for Pediatric Research were hand searched for RCTs (1992-1995). Subsequent publication was ascertained through a search of various electronic databases. Quality of abstracts and manuscripts was measured, and data were extracted using a structured form.

**Results:** A total of 264 (59.1%) of 447 abstracts were subsequently published. Almost 64% of RCTs that were subsequently published favored new therapy compared with 43.5% of studies that were never published ( $P < .001$ ). Mean effect size for published vs unpublished RCTs was 0.74 vs 0.05 ( $P < .001$ ). Median sample size was larger in published ( $n = 45$ ) vs unpublished ( $n = 34$ ) RCTs ( $P = .02$ ). Quality was significantly lower for abstracts vs published RCTs ( $P < .001$ ). For 5% of abstracts that were subsequently published, the conclusion regarding treatment efficacy changed.

**Conclusions:** Publication bias is a serious threat to assessing the effectiveness of interventions in child health, as little more than half of RCTs presented at a major scientific meeting are subsequently published. There is a need to institute an international registry of RCTs in children so that the totality of evidence can be accessed when assessing treatment effectiveness.

*Arch Pediatr Adolesc Med.* 2002;156:474-479

**P**UBLICATION BIAS, or the selective publication of studies that favor new therapies, poses a serious threat to validly assessing the effectiveness of new therapies.<sup>1</sup> To practice evidence-based medicine, the totality of evidence for and against a new therapy needs to be considered. The selective publication or submission of only those manuscripts that show a beneficial effect of therapy has important implications on the pool of evidence available on which to base decisions. This has implications for the individual clinician and for researchers who aim to synthesize existing evidence in the form of a systematic review.

In a recent meta-analysis evaluating the effectiveness of glucocorticoids for children with croup, Ausejo et al<sup>2</sup> demonstrated the challenge of selective publication. Formal statistical testing showed that studies with smaller sample sizes tended to show larger positive effects of an inter-

vention. One plausible explanation for this observation is publication bias.<sup>3</sup> This finding casts uncertainty on the degree of benefit conferred on patients with croup from glucocorticoids compared with placebo. McAuley and colleagues<sup>4</sup> showed that excluding the gray literature (ie, literature that is difficult to identify or retrieve, of which abstracts form the largest proportion at 61%) may yield larger estimates in meta-analyses by 15%.

### *For editorial comment see page 424*

Scientific meetings are often the first forum in which new research is shared with one's peers; hence, they represent a unique opportunity to examine the nature and existence of selective publication.<sup>5,6</sup> We undertook this study to (1) measure the percentage of randomized trials presented at a major pediatric sci-

*From the Department of Pediatrics, University of Alberta, and the Alberta Research Centre for Child Health Evidence, Edmonton (Drs Klassen and Craig and Mss Wiebe, Russell, Stevens, and Hartling); and the Thomas C. Chalmers Centre for Systematic Reviews, Children's Hospital of Eastern Ontario, University of Ottawa (Mr Moher).*

## MATERIALS AND METHODS

### SELECTION OF ABSTRACTS AND MANUSCRIPTS

Abstracts of randomized controlled trials (RCTs) were identified by hand-searching the proceedings from the Society for Pediatric Research (1992-1995). To determine whether an abstract was later published in manuscript form, the following databases were searched between February 1 and July 31, 2000: PubMed, EMBASE, Cochrane Library, CINAHL, Web of Science, Current Contents, and HEALTHSTAR. The databases were searched with the assistance of a medical librarian using the name of the primary author and key words found in the title. At least one common outcome must have been found in the abstract and the manuscript to be considered as the corresponding manuscript. If a relevant citation was not found in any of the databases, we assumed that the study had not been published in manuscript form.

Abstracts and manuscripts were included if they reported phase III RCTs with pediatric outcomes (studies reporting outcomes on pregnant women were excluded). Studies were excluded if they reported outcomes of nonrandomized treatment arms only.

### ASSESSMENT OF METHODOLOGICAL QUALITY

The quality, or internal validity, of the abstracts and manuscripts was assessed in 3 ways. First, study quality was scored according to a validated 5-point scale developed by Jadad et al.<sup>7</sup> This scale is composed of 5 questions to determine whether (1) the study was described as randomized, (2) an appropriate method of randomization was used (eg, table of random numbers), (3) the study was described as double blinded, (4) blinding was appropriate (eg, identical placebo), and (5) there was an adequate description of withdrawals and dropouts. A score of 5 indicates a high-quality study.<sup>7</sup> Second, concealment of allocation, or the method used to prevent foreknowledge of group assignment by the patients and the investigators, was assessed. Allocation concealment was rated as adequate (eg, centralized or pharmacy-controlled randomization), inadequate (ie, any procedure that is transparent before allocation, such as an open list of random numbers), or unclear or not reported.<sup>8</sup> Third, funding source was recorded as pharmaceutical, government, private, other, or unclear or not reported.

### DATA EXTRACTION

The following information was recorded for abstracts and manuscripts: journal citation, year of publication, diagnostic

category (as designated in the proceedings of the Society for Pediatric Research), trial design (ie, parallel or crossover), study type (ie, efficacy vs equivalency), stage of study (ie, preliminary or completed), pilot study vs full trial, number of withdrawals, and results.

The overall study conclusions (ie, favored the intervention or not) were determined by the authors' statements in their concluding remarks. If the study favored an intervention, the efficacious treatment was noted. Primary outcomes were determined in 1 of 3 ways: stated by the authors (19.9%), inferred by the extractors through the title or concluding remarks (32.2%), or randomly selected from probable primary outcomes using a computer-generated table of random numbers (47.9%).

Study type was defined according to the authors' statements with respect to the primary hypothesis. If the authors intended to demonstrate a significant difference between treatments, the study type was recorded as efficacy. Equivalence was deemed the study type when authors intended to show that there was no significant difference between treatments.

### STATISTICAL ANALYSIS

Data were entered into a database program (Access; Microsoft Corp, Redmond, Wash) and were analyzed using a statistical package (S-PLUS 2000; MathSoft Inc, Cambridge, Mass). Medians and interquartile ranges were used to describe nonparametric data, means and SDs were used for normal continuous data, and percentages and 95% confidence intervals (CIs) were used for dichotomous and categorical data. Odds ratios were converted into effect sizes using a method devised by Hasselblad and Hedges.<sup>9</sup> Percentages and their associated 95% CIs were also calculated for descriptive purposes and to quantify the change in variables from abstract to manuscript. To measure the association between different variables and publication status, the following tests were used: Pearson  $\chi^2$  tests for dichotomous and other categorical data, *t* tests (Welch modified for unequal variances) for normal continuous data, and Wilcoxon rank sum tests for otherwise nonparametric data. Logistic regression was used to evaluate predictors of publication while controlling for possible confounders such as overall study conclusions and study type. We expected that overall study conclusions would interact with study type. Thus, without the interaction term, study type might reduce the benefit attributed to the overall study conclusions. Log-rank tests were used to assess time to publication. Funnel plots (sample size vs effect size), the rank correlation test,<sup>10</sup> and weighted regression<sup>3</sup> were used to determine publication bias.

entific meeting that were subsequently published, (2) investigate factors associated with publication, and (3) describe the variables that changed from abstract to manuscript form.

## RESULTS

### SAMPLE

A total of 447 abstracts were identified; 264 (59.1%; 95% CI, 54.4%-63.7%) were found as published manuscripts.

**Table 1. Abstracts Published in Manuscript Form, by Year\***

Year	Abstracts, Total No.	Published, % (95% CI)
1992	95	57.9 (47.4-68.3)
1993	109	64.2 (54.8-73.7)
1994	128	60.9 (52.1-69.8)
1995	115	53.0 (43.5-62.6)
<b>Total</b>	<b>447</b>	<b>59.1 (54.4-63.7)</b>

\*CI indicates confidence interval.

**Table 2. Quality Variables by Document Type\***

Quality Measure	Abstracts		Manuscripts (n = 248)
	Not Subsequently Published (n = 183)	Subsequently Published (n = 264)	
Jadad score, median (interquartile range)	1 (1, 1)	1 (1, 1)	2 (1, 3)
Allocation concealment, % (95% CI)			
Adequate	0	0	31.9 (25.9-37.9)
Unclear	100	100	65.3 (59.2-71.4)
Inadequate	0	0	2.8 (0.6-5.1)
Funding source, % (95% CI)			
Government	2.7 (0.1-5.4)	4.9 (2.1-7.7)	13.7 (9.2-18.2)
Pharmacy	20.2 (14.1-26.3)	17.0 (12.3-21.8)	26.6 (20.9-32.3)
Private	2.2 (0-4.6)	3.0 (0.8-5.3)	19.4 (14.2-24.5)
Other	3.3 (0.4-6.1)	1.5 (0-3.2)	14.5 (9.9-19.1)
Unclear	71.6 (64.8-78.4)	73.5 (68.0-79.0)	25.8 (20.2-31.5)

\*CI indicates confidence interval. Difference between abstract groups: Jadad score,  $P = .8$ ; funding source,  $P = .44$ . Difference between manuscripts and all abstracts:  $P < .001$  for all 3 quality measures.

**Table 1** gives the number of abstracts identified and the percentage published in manuscript form, by year.

Sixteen manuscripts were excluded for the following reasons: not being RCTs ( $n = 10$ , although their abstracts claimed randomization), cluster designs ( $n = 2$ ), data authenticity questioned by the journal editors ( $n = 2$ ), unavailable methods ( $n = 1$ ), and study presented as a brief communication ( $n = 1$ ). The additional information gained by reading the manuscripts was not used to exclude abstracts because further information was not available for abstracts not subsequently published. The included manuscripts ( $n = 248$ ) were used exclusively in the change in variables (from abstract to manuscript form) analysis. The other analyses were based entirely on abstract data.

Abstracts were classified into 26 diagnostic categories. The most common categories were neonatology-general (24.6% of abstracts that remained unpublished and 23.1% of those that were subsequently published), neonatal pulmonology (14.8% of those unpublished and 12.9% of those published), neonatal nutrition and metabolism (12.6% of those unpublished and 11.4% of those published), and infectious diseases (7.1% of those unpublished and 10.6% of those published). The remaining categories were used for less than 10% of either type of abstract.

Manuscripts were published in 62 journals: 24.2% in *Journal of Pediatrics*, 20.6% in *Pediatrics*, and the remaining 55.2% in a variety of journals (with  $\leq 4.0\%$  in any one journal).

There were few abstracts on pilot studies (3.3% of unpublished abstracts and 2.3% of published abstracts) or studies in their preliminary stages (8.2% of unpublished abstracts and 7.6% of published abstracts). The number of withdrawals was reported in only 6.0% and 3.4% of unpublished and published abstracts, respectively. These variables were not considered in the logistic regression analysis.

#### QUALITY OF ABSTRACTS AND MANUSCRIPTS

**Table 2** summarizes study quality by document type. There were no differences in quality measures between

abstracts that were subsequently published and those that remained unpublished. Scores on the quality variables were significantly different for all abstracts compared with published manuscripts. Quality variables were not associated with overall study conclusions favoring the new therapy (Jadad score,  $P = .85$ ; funding source,  $P = .81$ ).

#### PREDICTORS OF PUBLICATION

**Table 3** provides the results of univariate analyses assessing predictors of publication. Variables associated with publication were study conclusions favoring treatment (63.5% vs 43.7%;  $P < .001$ ), sample size ( $n = 45$  vs 34;  $P = .02$ ), and measures of treatment effect. Measures of treatment effect were evaluated in 3 ways: odds ratios for dichotomous outcomes ( $P = .009$ ), standardized mean differences for continuous outcomes ( $P = .01$ ), and overall effect sizes (odds ratios and standardized mean differences combined) ( $P < .001$ ).

**Table 4** gives the results of logistic regression analysis of predictors of publication. The interaction between study type and overall study conclusions was significant. Successful equivalency studies, that is, studies in which the statistical results were nonsignificant, thereby agreeing with the study hypothesis, had the greatest odds of being published. Sample size was also significant, with the chances of publication increasing with larger sample size. Because the distribution of sample size was strongly skewed to the right, greater and greater increases are demanded to produce an equivalent predicting effect. This diminishing return is captured by a log-log transformation of the sample size variable. Effect size was not put into the model because its properties were closely associated with overall study conclusions.

#### VARIABLES THAT CHANGE FROM THE ABSTRACT TO THE PUBLISHED MANUSCRIPT

Two hundred forty-eight abstracts were paired with their subsequent manuscripts. Of these, 239 had conclusions favoring treatment; 2.1% changed from favoring treat-

**Table 3. Predictors of Publication: Univariate Results\***

Variable	Abstracts		P Value
	Not Subsequently Published	Subsequently Published	
<b>Dichotomous Variables</b>			
Overall study conclusions favoring treatment	43.7 (35.9-51.5) [n = 167]	63.5 (57.4-69.6) [n = 255]	<.001
Equivalency studies (vs efficacy)	10.9 (6.1-15.7) [n = 183]	16.3 (11.6-20.9) [n = 264]	.13
Parallel study design (vs crossover)	86.3 (81.1-91.6) [n = 183]	87.8 (83.7-92.0) [n = 263]	.78
Neonatal classification (vs pediatric)	59.0 (51.6-66.4) [n = 183]	57.6 (51.4-63.7) [n = 264]	.84
<b>Categorical Variable</b>			
Year			.35
1992	21.9 (15.6-28.1)	20.8 (15.7-25.9)	
1993	21.3 (15.1-27.5)	26.5 (21.0-32.0)	
1994	27.3 (20.6-34.1)	29.5 (23.9-35.2)	
1995	29.5 (22.6-36.4) [n = 183]	23.1 (17.8-28.4) [n = 264]	
<b>Discrete Nonparametric Variables</b>			
Sample size, median (interquartile range)	34 (17-78) [n = 147]	45 (20-115.5) [n = 216]	.02
<b>Continuous Normal Variables</b>			
OR (for dichotomous outcomes), mean (95% CI)†	0.647 (0.415 to 1.008) [n = 38]	0.305 (0.217 to 0.429) [n = 60]	.009
SMD (for continuous outcomes), mean (95% CI)†	0.114 (-0.468 to 0.696) [n = 44]	-0.814 (-1.197 to -0.432) [n = 63]	.01
Effect sizes (for all outcomes), mean (95% CI)†‡	-0.05 (-0.382 to 0.282) [n = 82]	-0.737 (-0.952 to -0.521) [n = 123]	<.001

\*Data are given as percentage (95% confidence interval [CI]) except where indicated otherwise. OR indicates odds ratio; SMD, standardized mean difference.

†Beneficial treatment effect indicated by ORs <1, negative SMDs, and negative effect sizes.

‡The ORs and SMDs are combined into overall effect sizes.<sup>9</sup>

**Table 4. Predictors of Publication: Results From Logistic Regression (n = 343)\***

Variable	Odds Ratio (95% CI)
Unsuccessful efficacy study	1.000 (Referent)
Unsuccessful equivalency study	2.532 (0.555-11.494)
Unsuccessful efficacy study	4.032 (2.421-6.757)
Successful equivalency study	6.173 (2.646-14.286)
Sample size (for every log-log unit increase)†	2.681 (1.332-5.405)

\*CI indicates confidence interval.

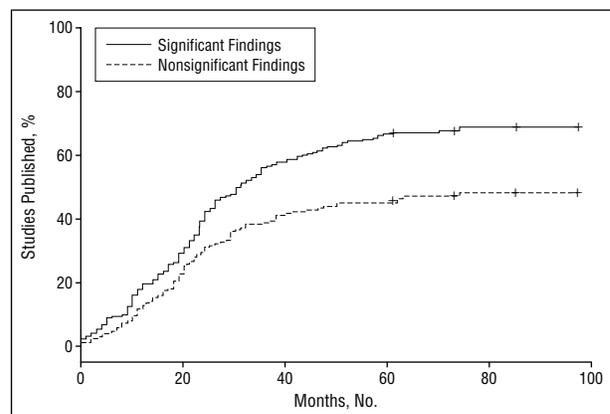
†Right-skewed independent variable normalized to strengthen model fit.

ment in the abstract to not favoring treatment in the manuscript, whereas 2.9% changed from not favoring treatment in the abstract to favoring treatment in the manuscript.

Most measures of treatment effect changed slightly from the abstract to the manuscript. Thirty-eight percent and 45.1% decreased in log odds ratios and standardized mean differences, respectively, whereas 44.0% and 37.3% increased. Overall significance in effect size was also evaluated ( $P \leq .05$ ). In 4.0% of the cases, abstract outcomes were significant but changed to nonsignificant in manuscript form; 8.9% changed from nonsignificant to significant. In 10.8% of studies, the sample size decreased from the abstract to the manuscript, whereas in 48.3% the sample size increased.

#### TIME TO PUBLICATION

**Figure 1** shows the time to publication by overall study conclusions. The probability of publication plateaued after approximately 60 months. After 5 years, 34.2% of stud-



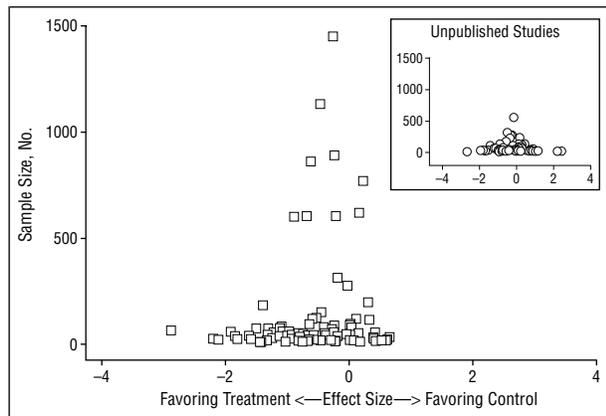
**Figure 1.** Time to publication by significant and nonsignificant findings.

ies with significant findings remained unpublished compared with 55.7% of studies with nonsignificant results.

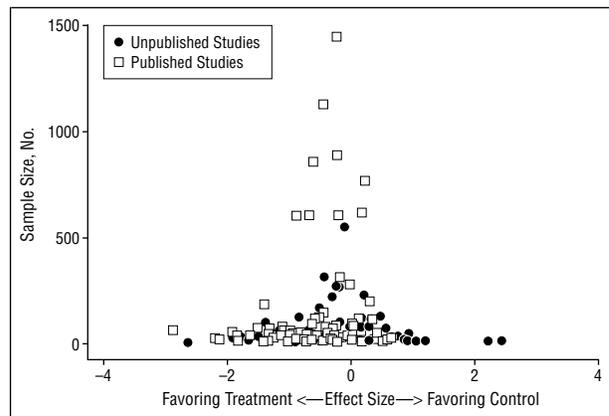
Of studies that were subsequently published, abstracts with Jadad scores of 1 or less were published on average 24.7 months after the abstract appeared in the proceedings of the Society for Pediatric Research compared with 21.3 months for those with scores greater than 1 ( $P = .11$ ). There was a significant difference between abstracts that did not report a funding source (25.1 months) and those that did (20.9 months) ( $P = .049$ ). Favoring the new therapy was not associated with time to publication ( $P = .74$ ).

#### PUBLICATION BIAS

**Figure 2** and **Figure 3** show funnel plots for published and unpublished studies. In the absence of publication bias, the funnel plot should resemble an inverted funnel. Smaller studies naturally (or statistically)



**Figure 2.** Funnel plot of published studies. Outliers: -9.4, 25; -5.1, 14; -1.4, 6794; -0.6, 2342; -0.2, 5699; and 0.1, 2416.



**Figure 3.** Funnel plot of all studies. Outliers: -9.4, 25; -5.1, 14; -1.4, 6794; -0.6, 2342; -0.2, 5699; 0.1, 2416; and 10.8, 7.

give a wide range of effect sizes; they are less stable because they contain less information. Thus, at the base of the inverted funnel, one would expect to see the widest spread. At the top of the plot, the effect sizes should converge at a point, if there exists a trial that is sufficiently large to be relatively stable. The funnel plot of published studies (Figure 2) and statistical tests for publication bias show that small unsuccessful studies are less likely to be published (rank correlation test,  $P = .03$ ; weighted regression,  $P = .02$ ); the base of the inverted funnel is not balanced, resulting in funnel plot asymmetry. No asymmetry was observed in the sample of studies that were not subsequently published (rank correlation test,  $P = .43$ ; weighted regression,  $P = .62$ ) (Figure 3). All studies combined (Figure 3) suggest evidence of publication bias that predates submission to the pediatric meeting (rank correlation test,  $P = .07$ ; weighted regression,  $P = .02$ ). Missing studies favoring the control groups are indicated by the asymmetry of the plot. Publication bias may include nonsubmission of studies to the meetings or to journals (submission bias) and rejection of submissions by editorial staff (publication bias).

#### COMMENT

The results of this study demonstrate that a major factor predicting publication after presentation of an RCT at a scientific meeting is a positive outcome favoring the new therapy. The larger the treatment effect, the more likely that a study will be published. This is the first time to our knowledge that this has been shown in the area of child health but is consistent with results of studies in other areas.<sup>1,11,12</sup> One implication of this finding is that by using only the published literature to examine the effectiveness of a new therapy, one could get an inflated view of the benefit of a new therapy. Even more worrisome is to conclude that a treatment is effective when if the totality of evidence had been identified (published and unpublished) it may indeed not be better than existing treatments.

It is not known how many RCTs performed in children are never submitted to a scientific meeting. Although the collection of RCTs presented at scientific meetings does not capture the entire universe of RCTs performed in children, it does provide a useful filter to examine the

question of publication bias. The funnel plot of all studies submitted to the Society for Pediatric Research meeting indicates that submission bias exists (Figure 3), albeit this bias is less pronounced than in studies that are published as full manuscripts (Figure 2).

The results of our analyses indicate that studies with negative findings are less likely to be published. Stern and Simes<sup>13</sup> examined clinical trials approved by a research ethics board for 10 years and found that the hazard ratio for studies with positive vs negative findings being published was 3.13 ( $P < .001$ ). Scherer and colleagues<sup>14</sup> conducted a meta-analysis on studies that documented the proportion of studies published after presentation at a scientific meeting and found an overall estimate of 51%, which is similar to our estimate of 59.1%. Two of these studies, in the areas of ophthalmology and perinatology, were confined to RCTs, and reported publication rates were 61% and 36%, respectively.

Several options have been proposed to minimize publication bias. An international registry of RCTs has been suggested<sup>1,15,16</sup> whereby all trials would be registered at their inception. A standard method of identifying trials for registration would be through ethics review boards, as all trials require ethics approval before beginning. Recently, an amnesty for authors of unpublished trials was initiated. One hundred medical journals solicited information on unpublished research by requesting authors to complete an unreported trial registration form.<sup>17</sup> This initiative had modest results, with 165 trials involving 32 000 participants being registered within the first year.<sup>18</sup>

Until publication bias can be fully prevented, there is a need to quantify its existence and to interpret study findings accordingly. Methods have been developed for the detection of publication bias, but they lack sufficient statistical power in meta-analyses with fewer than 10 trials.<sup>19</sup> In addition, their performance varies with factors such as sample size, number of trials, and size of treatment effect.<sup>20,21</sup>

Publication bias may be more frequent in child health as sample sizes are smaller than those in the general literature. Smaller sample sizes mean more statistical fluctuation, hence the probability that small studies may show large treatment effects. Small studies with large treatment effects that favor new therapies are more likely to

### What This Study Adds

#### What Is Already Known and Why This Study Needed to Be Done

- Publication bias is a serious threat to assessing new therapies
- The probability of bias increases if the study demonstrates greater benefit for the new therapy, hence seriously undermining our confidence in the published record of evidence as a true estimate of benefit

#### What This Study Adds to Medical Information and Its Implications

- This study demonstrates that publication bias is also a significant issue in child health as 40% of trials presented at a scientific meeting are not published
- When synthesizing therapy evidence, it is critical to assess for the possibility of publication bias
- Prevention of publication bias could occur if all randomized trials performed in children were mandated to undergo international registration at the time of ethics approval

be published than those that do not favor new therapies. It would be prudent for researchers in child health to move toward larger multicenter trials whenever possible to avoid this "small study effect."

In a survey of editors, 30% of them said they would not publish a meta-analysis that included unpublished research.<sup>22</sup> The meta-analyst is caught in a dilemma, as excluding abstracts or other unpublished RCTs means the meta-analysis may be biased toward favoring treatment. McAuley et al<sup>4</sup> demonstrated that the average quality of abstracts was lower than full manuscript reports of RCTs. They could not determine whether this was because lower-quality RCTs were not published or whether, as we demonstrate, the lower-quality assessments are because of limited information contained in an abstract. In our study, most abstracts (76%) had a Jadad score of 1. We know that with published manuscripts, RCTs that score less than 2 may have treatment effects 30% larger than high-quality studies (score >2).<sup>23</sup> The lack of complete reporting holds true for allocation concealment, as 100% of the abstracts had unclear allocation concealment. A structured reporting system for abstracts of RCTs that would require the author to document these key areas would help solve this problem without greatly increasing the length requirements of abstracts. This has worked well for reporting of RCTs in journals, and the adoption of these guidelines by journals has improved their quality of reporting.<sup>22,24</sup> Another solution would be an international registry of trials that would allow meta-analysts to access longer and more detailed versions of the RCT study report in a database.

Despite what has been known about publication bias for 40 years,<sup>25</sup> it remains a threat to assessing the validity of therapy in child health. It is urgent that funding agencies, governmental bodies, health administrators, researchers, and clinicians join together to eliminate this important and serious bias.

Accepted for publication January 10, 2002.

Corresponding author and reprints: Terry P. Klassen, MD, Department of Pediatrics, University of Alberta, 2C3.67 Walter C. Mackenzie Centre, Edmonton, Alberta, Canada T6G 2B7 (e-mail: terry.klassen@ualberta.ca).

### REFERENCES

1. Simes RJ. Publication bias: the case for an international registry of clinical trials. *J Clin Oncol.* 1986;4:1529-1541.
2. Ausejo M, Saenz A, Pham B, et al. The effectiveness of glucocorticoids in treating croup. *BMJ.* 1999;319:595-600.
3. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple graphical test. *BMJ.* 1997;315:629-634.
4. McAuley L, Pham B, Tugwell P, Moher D. Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses? *Lancet.* 2000;356:1228-1231.
5. Evers JLH. Publication bias in reproductive research. *Hum Reprod.* 2000;15:2063-2066.
6. Kiroff GK. Publication bias in presentations to annual scientific congress. *Aust N Z J Surg.* 2001;71:167-171.
7. Jadad AR, Moore RA, Carrol D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials.* 1996;17:1-12.
8. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. *JAMA.* 1995;273:408-412.
9. Hasselblad V, Hedges LV. Meta-analysis of screening and diagnostic tests. *Psychol Bull.* 1995;117:167-178.
10. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics.* 1994;50:1088-1101.
11. Dickersin K, Min Y-I, Meinert CL. Factors influencing publication of research results. *JAMA.* 1992;267:374-378.
12. Easterbrook PJ, Berlin JA, Gøtzsche PC, Matthews DR. Publication bias in clinical research. *Lancet.* 1991;337:867-872.
13. Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research papers. *BMJ.* 1997;315:640-645.
14. Scherer RW, Dickersin K, Langenberg P. Full publication of results initially presented in abstracts: a meta-analysis. *JAMA.* 1994;272:158-162.
15. Moher D. Clinical-trial registration: a call for its implementation in Canada. *CMAJ.* 1993;149:1657-1658.
16. Horton R, Smith R. Time to register randomized trials. *Lancet.* 1999;354:1138-1139.
17. Roberts I, Hoey J. An amnesty for unpublished trials [editorial]. *CMAJ.* 1997;157:1548.
18. Roberts I. An amnesty for unpublished trials: one year on, many trials are unregistered and the amnesty remains open [editorial]. *BMJ.* 1998;317:763-764.
19. Sutton AJ, Duval SJ, Tweedie RL, Abrams KR, Jones DR. Empirical assessment of effect of publication bias on meta-analyses. *BMJ.* 2000;320:1574-1577.
20. Sterne JAC, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol.* 2000;53:1119-1129.
21. Macaskill P, Walter SD, Irwig L. A comparison of methods to detect publication bias in meta-analysis. *Stat Med.* 2001;20:641-654.
22. Cook DJ, Guyatt GH, Ryan G. Should unpublished data be included in meta-analyses? *JAMA.* 1993;269:2749-2753.
23. Moher D, Pham B, Jones A, et al. Does the quality of report of randomized trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet.* 1998;352:609-613.
24. Moher D, Jones A, Lepage L, for the CONSORT Group. Use of the CONSORT statement and quality of reports of randomized trials: a comparative before-and-after evaluation. *JAMA.* 2000;285:1992-1995.
25. Sterling TD. Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa. *J Am Stat Assoc.* 1959;54:30-34.