BMC Medical Research Methodology

BioMed Central

Research article

Meta-analytic methods for pooling rates when follow-up duration varies: a case study

James P Guevara*[1], Jesse A Berlin[2] and Fredric M Wolf[3]

Address: [1]Department of Pediatrics, The Children's Hospital of Philadelphia and the University of Pennsylvania School of Medicine, 3535 Market St, Room 1531, Philadelphia, PA 19104, USA, [2]Department of Biostatistics and Epidemiology, Center for Clinical Epidemiology and Biostatistics, Center for Education and Research on Therapeutics, University of Pennsylvania School of Medicine, Blockley Hall, Room 611, Philadelphia, PA 19104, USA and [3]Department of Medical Education and Biomedical Informatics, University of Washington School of Medicine, 1959 NE Pacific St, Room E-312, Seattle, WA 98195, USA

Email: James P Guevara* - guevara@email.chop.edu; Jesse A Berlin - jberlin@cceb.upenn.edu; Fredric M Wolf - wolf@u.washington.edu

* Corresponding author

Published: 12 July 2004

BMC Medical Research Methodology 2004, **4**:17 doi:10.1186/1471-2288-4-17

Received: 19 March 2004

Accepted: 12 July 2004

This article is available from: http://www.biomedcentral.com/1471-2288/4/17

Abstract

Background: Meta-analysis can be used to pool rate measures across studies, but challenges arise when follow-up duration varies. Our objective was to compare different statistical approaches for pooling count data of varying follow-up times in terms of estimates of effect, precision, and clinical interpretability.

Methods: We examined data from a published Cochrane Review of asthma self-management education in children. We selected two rate measures with the largest number of contributing studies: school absences and emergency room (ER) visits. We estimated fixed- and random-effects standardized weighted mean differences (SMD), stratified incidence rate differences (IRD), and stratified incidence rate ratios (IRR). We also fit Poisson regression models, which allowed for further adjustment for clustering by study.

Results: For both outcomes, all methods gave qualitatively similar estimates of effect in favor of the intervention. For school absences, SMD showed modest results in favor of the intervention (SMD -0.14, 95% CI -0.23 to -0.04). IRD implied that the intervention reduced school absences by 1.8 days per year (IRD -0.15 days/child-month, 95% CI -0.19 to -0.11), while IRR suggested a 14% reduction in absences (IRR 0.86, 95% CI 0.83 to 0.90). For ER visits, SMD showed a modest benefit in favor of the intervention (SMD -0.27, 95% CI: -0.45 to -0.09). IRD implied that the intervention reduced ER visits by 1 visit every 2 years (IRD -0.04 visits/child-month, 95% CI: -0.05 to -0.03), while IRR suggested a 34% reduction in ER visits (IRR 0.66, 95% CI 0.59 to 0.74). In Poisson models, adjustment for clustering lowered the precision of the estimates relative to stratified IRR results. For ER visits but not school absences, failure to incorporate study indicators resulted in a different estimate of effect (unadjusted IRR 0.77, 95% CI 0.59 to 0.99).

Conclusions: Choice of method among the ones presented had little effect on inference but affected the clinical interpretability of the findings. Incidence rate methods gave more clinically interpretable results than SMD. Poisson regression allowed for further adjustment for heterogeneity across studies. These data suggest that analysts who want to improve the clinical interpretability of their findings should consider incidence rate methods.

Background

Meta-analysis has become recognized as an objective means of summarizing evidence from disparate clinical trials [1]. It is particularly useful when the trials are small and the data are conflicting. Meta-analysis incorporates statistical approaches to pool aggregate data from clinical trials into a summary effect measure [2]. This measure then reflects the effect of an intervention on average across all studies. However, meta-analysis is limited by inclusion of poor quality trials that are prone to report biased findings and exclusion of unpublished trials that do not report findings. Methods for assessing the effect of these limitations on summary measures have been developed and are available [3-5].

At times, data from clinical trials may conform to continuous rate measures (events per person-time) in which the numerator represents a count of total events "x" and the denominator represents a given time duration multiplied by the number of subjects, e.g. health care visits per person-year. Data such as these are being reported more frequently in clinical trials as evidenced by inclusion of rate measures in recent Cochrane Systematic Reviews [6-9]. If the reported length of follow-up is the same across studies, e.g. 12 months, then meta-analysis might involve pooling the weighted within-study differences in the mean number of events per person between intervention and control groups, a method we will call the weighted mean difference (WMD) [10]. The interpretation is straightforward and reflects the change in "x" per unit time. However, if the reported length of follow-up from various studies is different, e.g. 6 months versus 1 year, then meta-analysis could involve the conversion of the study differences into a common metric prior to pooling. This is often accomplished by dividing the per study differences between groups by the pooled standard deviation, a procedure known as the standardized weighted mean difference (SMD) [10]. This method is robust to assumptions of varying follow-up time. However, the interpretation is more difficult, since it reflects the difference between intervention and control groups measured in standard deviation units rather than natural time units.

In this paper, we examined data from a recently published Cochrane Systematic Review that included continuous rate measures as outcomes. We compared different statistical approaches to pooling continuous rate measures when they were reported with varying follow-up time. Specifically, we examined the SMD, considered the standard approach, to two alternative methods, incidence rate differences and incidence rate ratios. We examined the results from the different approaches in terms of the point estimates of treatment effect, their precision, and clinical interpretability. We are unaware of previously published studies that have attempted to address this problem.

Methods

Data were taken from a recently published Cochrane systematic review on the effects of asthma self-management education in children [11]. We selected the two outcomes involving continuous rate measures with the greatest number of contributing studies: days of school absence and emergency room (ER) visits. Our goal was to compare the standardized weighted mean difference with two alternative statistical approaches to pooling rate data, incidence rate differences and incidence rate ratios.

The standardized weighted mean difference (SMD) represents a weighted average of the per study difference in mean events per person between treatment and control groups. We first calculated standardized effect sizes for each study by subtracting the reported mean number of events in the control group from the reported mean number of events in the treatment group and dividing by the pooled standard deviation [10]. The per study standardized effect sizes were then combined using both fixed- and random-effects models [12,13]. The fixed-effects model is essentially a weighted average of the study-specific results in which the weight for each study is proportional to the inverse of the variance of the study-specific SMD. The random-effects model allows for variability among studies in the SMD by incorporating a term for the among-study variability into the weights. Fixed- and random-effects models will generally agree when there is little heterogeneity among studies.

To estimate stratified incidence rate differences (IRD) and stratified incidence rate ratios (IRR), we calculated incidence rates taking time explicitly into account. For each study, we knew the mean number of events (days absent or emergency room visits) and the number of months of observation according to the reported study design. We multiplied the mean by the sample size for each treatment arm to get the total number of events observed in each arm, e.g. the total number of days absent for all participants in the control group. We rounded this to the nearest whole number of events. To obtain the total person-time of follow-up, we assumed that there was no loss to follow-up during the study, i.e. all participants were observed for the entire length of the study. We multiplied the number of months of follow-up by the sample size for each arm to obtain the total number of person-months of follow-up. The study-specific rate of events per person-month for each arm was then the total number of events (days absent or emergency room visits) divided by the total number of person-months of observation for each arm.

The analysis of the rates used stratified IRD and IRR methods estimated in STATA (version 7). To obtain a summary stratified IRD, we used a program, co-written by one of us (JAB) to implement a fixed-effects Mantel-Haenszel (M-

H) procedure in STATA. Specifically, the program produced the estimates of IRD and its variance described in Rothman and Greenland's textbook [14]. We also utilized an inverse-variance weighted average approach to estimate a random-effects models by first using STATA's "ird" command, saving those study-specific results, then using the STATA command "meta" to compute the weighted average IRDs [13].

To obtain a summary stratified IRR, we used a fixed-effects M-H type procedure as implemented in the "ir" command in STATA, which should give results similar to fitting a Poisson regression model with indicator variables for "study." This M-H approach produces a summary estimate stratified on study. To take study-to-study variability into account, we also fit Poisson regression models allowing for clustering of the data by study, both with and without study indicator variables. The inclusion of indicator variables forces the comparison between treatments to be made within study, thereby mimicking the stratified analysis. In STATA, we also fit Poisson regression models using the "cluster" option, which uses a robust (Huber-White "sandwich") estimator of the variance [15]. The intent of fitting these models that allowed for clustering was to inflate the variance estimates to allow for among-study variability, and (as will be demonstrated) would *not* affect the point estimates of treatment effect.

Our interest was in comparing the qualitative and, where possible, the quantitative results across the different methods. We were interested in differences in inference that could be made from the various models, which integrate information about the point estimates of treatment effects and the precision of their estimation but may vary in their assumptions. We also compared conclusions as to the heterogeneity of effects across studies. The methods based on weighted averages use a test of heterogeneity similar in principle to the Cochrane Q-statistic. The test for heterogeneity in the Poisson regression models is based on the interactions between the treatment variable and the study indicator variables. Most importantly, we were concerned with the clinical interpretability of the results. All p-values reported are two-sided and all confidence intervals are calculated at the 95% level.

Results

We illustrate the use of SMD, IRD, and IRR methods for pooling continuous rate measures using data from a published Cochrane systematic review and meta-analysis that examined the effect of self-management education on morbidity and health services outcomes in children and adolescents with asthma [11]. The meta-analysis included 32 separate trials, involving 3706 children and adolescents aged 2 to 18 years. The majority were small, randomized controlled trials that enrolled children with severe asthma. We abstracted data on two outcomes–days of school absence and ER visits – from the original published study. For each outcome, we abstracted the reported mean number of events, standard deviation, sample size, and observation time in months for treatment and control groups. We contacted study authors to identify missing data from published reports. If appropriate measures of variance were not reported nor obtained by author contact, we imputed pooled standard deviations using a conservative approach given the t-statistic or the p-value if the t-statistic was not reported [16].

Table 1 lists the treatment and control group sizes, mean number of events, standard deviations, rates (events/person-month), duration of follow-up, and standardized effect size for each of the 16 trials contributing data on school absences. Sample sizes ranged from 19 to 404 participants, and the duration of observation varied widely from 1 to 12 months. Most of the trials favored the treatment arm, i.e. negative effect sizes implied a reduction in school absences. However, larger studies tended to have standardized effect size estimates closer to the null.

Similarly, table 2 lists the treatment and control group sizes, mean number of events, standard deviations, rates (events/person-month), duration of follow-up, and standardized effect size for each of the 12 trials contributing data on ER visits. Again, sample sizes ranged from a low of 14 to a high of 232, but the duration of follow-up was more homogenous with most trials reporting 12 months of observation. Again, most trials favored the treatment arm. Similar to school absences, larger studies tended to have standardized effect size estimates closer to the null.

Table 3 presents the summary outcome measures for school absences. Effect sizes from the 3 methods gave qualitatively similar conclusions and suggest that treatment reduces school absences. Both fixed- and random-effects SMD gave identical estimates, since there was little to no statistical heterogeneity present ($p = 0.61$). IRD methods gave clinically interpretable results on the absolute scale. The fixed-effects results suggest that treatment results in an average reduction of 0.15 school absences per child per month (1.8 absences per year). Random-effects estimates were consistent with the fixed-effects results but with wider confidence intervals. IRR methods gave clinically interpretable results on the relative scale. These results suggest that treatment results in a 14% reduction in school absences. IRR estimates obtained using Poisson regression with Huber-White sandwich estimators gave a more conservative estimate than IRR estimates obtained using M-H procedures. The IRR estimate obtained without study indicators was similar to the IRR estimate with study indicators, suggesting that confounding by study was not

Table 1: Characteristics of Studies Reporting on School Absences.*

Study	Intervention Group			Control Group			Duration (Months)	Standardized Effect Size**
	N	Mean ± SD	Rate	N	Mean ± SD	Rate		
Charlton	42	2.10 ± 11.40	0.18	37	4.70 ± 15.50	0.39	12	-0.19
Christiansen	27	2.39 ± 2.90	0.20	15	2.98 ± 3.29	0.25	12	-0.19
Colland	45	0.98 ± 1.56	0.16	34	0.53 ± 1.08	0.09	6	0.32
Dahl	9	0.80 ± 0.32	0.8	10	0.90 ± 0.32	0.9	1	-0.30
Deaves	32	3.69 ± 4.80	0.31	31	5.19 ± 4.80	0.43	12	-0.31
Evans	117	19.40 ± 13.90	1.62	87	19.70 ± 12.60	1.64	12	-0.02
Fireman	13	0.50 ± 5.06	0.04	13	4.60 ± 5.06	0.38	12	-0.78
Hill	211	5.43 ± 4.07	1.36	193	6.23 ± 4.72	1.56	4	-0.18
Hughes	44	10.70 ± 6.90	0.89	45	16.00 ± 15.40	1.33	12	-0.44
Mitchell	133	7.92 ± 16.48	1.32	126	8.48 ± 26.69	1.41	6	-0.03
Perrin	29	0.24 ± 0.90	0.24	27	0.22 ± 1.00	0.22	1	0.02
Persaud	18	6.40 ± 4.60	1.28	18	7.60 ± 5.30	1.52	5	-0.24
Rubin	29	11.90 ± 7.80	0.99	25	15.40 ± 15.00	1.28	12	-0.30
Talabere	25	1.36 ± 2.52	0.45	25	2.60 ± 3.75	0.87	3	-0.38
Toelle	63	2.62 ± 3.28	0.44	51	2.67 ± 3.21	0.45	6	-0.02
Wilson	30	0.80 ± 2.29	0.80	29	1.40 ± 3.23	1.40	1	-0.21

* N refers to the sample size, Mean ± SD refers to the mean number of events ± standard deviation, and rate refers to the total events per person-month. ** Standardized effect size was calculated for each study by subtracting control group mean from intervention group mean and dividing by the pooled SD.

Table 2: Characteristics of Studies Reporting on Emergency Room Visits.*

Study	Intervention Group			Control Group			Duration (Months)	Standardized Effect Size**
	N	Mean ± SD	Rate	N	Mean ± SD	Rate		
Alexander	11	0.60 ± 0.90	0.05	10	2.40 ± 2.10	0.20	12	-1.09
Christiansen	27	0.30 ± 1.20	0.03	15	0.20 ± 0.43	0.02	12	0.10
Clark	159	1.72 ± 4.20	0.14	73	2.49 ± 6.26	0.21	12	-0.16
Fireman	13	0.08 ± 1.14	0.01	13	1.00 ± 1.14	0.08	12	-0.78
Hughes	44	0.45 ± 1.05	0.04	45	0.60 ± 1.05	0.05	12	-0.14
Lewis	48	2.30 ± 2.98	0.19	28	3.71 ± 2.98	0.31	12	-0.47
McNabb	7	1.90 ± 4.72	0.16	7	7.40 ± 4.72	0.62	12	-1.09
Persaud	18	0.27 ± 0.57	0.05	18	1.00 ± 1.20	0.20	5	-0.76
Ronchetti	114	0.07 ± 0.32	0.01	95	0.23 ± 0.78	0.02	12	-0.28
Shields	101	0.54 ± 1.68	0.05	104	0.38 ± 1.68	0.03	12	0.09
Talabere	25	0.44 ± 0.77	0.15	25	1.08 ± 1.32	0.36	3	-0.58
Toelle	63	1.51 ± 2.31	0.25	51	1.67 ± 2.40	0.28	6	-0.07

* N refers to the sample size, Mean ± SD refers to the mean number of events ± standard deviation, and rate refers to the total events per person-month. ** Standardized effect size was calculated for each study by subtracting control group mean from intervention group mean and dividing by the pooled SD.

present for this outcome (see the Appendix for further discussion of this point). Heterogeneity was statistically detected when data were pooled using IRD ($p < 0.001$) and IRR methods ($p < 0.001$) but not SMD, suggesting that treatment effects varied across studies when assessed in terms of rates, but not when assessed in terms of standard deviation units.

Table 4 presents summary outcomes measures for ER visits. Results were again qualitatively similar regardless of method and suggest that treatment reduces ER visits. Random-effects SMD gave a more conservative estimate with wider confidence intervals than the corresponding fixed effects SMD, due to heterogeneity in effects across the studies ($p = 0.05$). IRD methods gave clinically interpretable results on the absolute scale: treatment results in an average reduction of 0.04 ER visits per child per month

Table 3: Summary Outcome Measures for Days of School Absence.

Measure	Effect Size	Confidence Interval	Effect Size P-value	Homogeneity Test P-value
SMD[a]				
Fixed-effects	-0.14	-0.23, -0.04	0.006	0.61
Random-effects	-0.14	-0.23, -0.04	0.006	0.61
IRD[b]				
Fixed-effects M-H	-0.15	-0.19, -0.11	<0.001	<0.001
Random-effects	-0.17	-0.25, -0.08	<0.001	<0.001
IRR[c]				
Fixed-effects M-H	0.86	0.83, 0.90	<0.001	<0.001
PR + study indicators	0.86	0.77, 0.97	0.011	<0.001
PR - study indicators	0.86	0.75, 0.99	0.044	N/A

[a] SMD refers to standardized mean difference and was obtained using both fixed effects and random effects models. [b] IRD refers to the incidence rate difference, and was obtained using a Mantel-Haenszel procedure to estimate a fixed-effects model and an inverse-variance method to estimate a random-effects model. [c] IRR refers to the incidence rate ratio and was obtained using Mantel-Haenszel procedure to estimate a fixed effects model and Poisson regression models with Huber-White sandwich estimators with and without study indicators which is equivalent to a random-effects model.

Table 4: Summary Outcome Measures for Emergency Room Visits.

Measure	Effect Size	Confidence Interval	Effect Size P-value	Homogeneity Test P-value
SMD[a]				
Fixed-effects	-0.21	-0.33, -0.09	<0.001	0.05
Random-effects	-0.27	-0.45, -0.09	0.003	0.05
IRD[b]				
Fixed-effects M-H	-0.04	-0.05, -0.03	<0.001	<0.001
Random-effects	-0.05	-0.08, -0.03	<0.001	<0.001
IRR[c]				
Fixed-effects M-H	0.66	0.59, 0.74	<0.001	<0.001
PR + study indicators	0.66	0.54, 0.81	<0.001	<0.001
PR – study indicators	0.77	0.59, 0.99	0.039	N/A

[a] SMD refers to standardized mean difference and was obtained using both fixed effects and random effects models. [b] IRD refers to the incidence rate difference, and was obtained using a Mantel-Haenszel procedure to estimate a fixed-effects model and an inverse-variance method to estimate a random-effects model. [c] IRR refers to the incidence rate ratio and was obtained using Mantel-Haenszel procedure to estimate a fixed effects model and Poisson regression models with Huber-White sandwich estimators with and without study indicators which is equivalent to a random-effects model.

(one ER visit every other year). The estimate obtained by the random-effects model was consistent with that from the fixed-effects model but with wider confidence intervals. IRR methods gave clinically interpretable results on the relative scale: treatment results in a 23 to 34% reduction in ER visits. IRR estimates obtained using Poisson regression with Huber-White sandwich estimators gave a more conservative estimate than IRR estimates obtained using M-H procedures. The IRR estimate obtained without study indicators was closer to the null than the IRR estimate with study indicators, suggesting that confounding by study was present for this outcome (see Appendix). Heterogeneity was statistically present in IRD ($p < 0.001$) and IRR ($p < 0.001$) methods as well as for SMD for this outcome, suggesting that treatment effects varied across studies.

Discussion

This paper presented three statistical methods of pooling continuous rate measures in which the denominator reflects varying duration of observation. All methods were fairly easy to implement using standard statistical software. Results were statistically consistent regardless of the method employed and suggested a significant treatment effect on average. All methods allowed for explicit adjustment for individual studies. Failure to take stratification by study into account, as illustrated in the Poisson models without study indicators, resulted in a different estimate

for one outcome, ER visits, but not the other, school absences.

IRD methods gave clinically interpretable results on an absolute scale. These results suggest that treatment results in an average reduction of 0.15 school absences per person-month or roughly 2 days per person-year. These results also suggest that treatment results in an average of 0.04 fewer ER visits per person-month or roughly 1 fewer visit per person every 2 years. IRR methods gave clinically interpretable results on a relative scale. These results suggest that treatment results in a 14% reduction in school absences and a 34% reduction in ER visits.

The SMD results were not immediately clinically interpretable. On a standard deviation scale, these results suggest that treatment results in a modest reduction in school absences and ER visits. Conversion back to the original scale would allow for more clinically interpretable results but would require making an assumption about the size of the standard deviation and the event rate in the control group across studies. For standard deviations, it is not clear whether one should use a study-specific estimate of the standard deviation or an estimate pooled across studies. Additionally, the data can be skewed, in which case mean events might not appropriately represent the central tendency of the data.

Heterogeneity was statistically present for both outcomes, suggesting variability in treatment effects across studies when incidence rate-based methods were used, and for ED visits but not school absences when SMD was used. It should be kept in mind that, although all of these analyses are attempting to address the same underlying substantive question (i.e., whether asthma education "works"), the SMD analyses address this question on a fundamentally different scale by converting measurements into standard deviation units. This difference in scale could well account for the different results of the heterogeneity tests.

Another alternative that we tried but abandoned because of its non-standard nature was simply to convert the time units from the various studies into a common scale and pool the data using WMD. We found (data not shown) slight but noticeable differences depending on whether we multiplied up for the shorter studies or down for the longer studies to achieve the common scale. For example, studies with 6-month follow-up and 12-month follow-up could be put on a common scale, by either multiplying the 6-month study means and standard deviations by 2 or dividing the 12-month study means and standard deviations by 2. These different approaches changed the per-study weights and produced slight differences in summary measures. We believe that the fundamental problem with this approach is that it rests on the assumption that the event rates stay constant over the entire time period of observations. This is also true for the rate models we did use, but unlike those models, multiplying up essentially imputes data beyond the actual period of observation. This has implications not only for the mean number of events, but possibly also for the variance estimates. For these reasons, we chose not to consider this approach any further.

There are limitations to these findings. First, we explored differences in the three approaches using only data from a single systematic review. However, the outcomes we chose had a sufficient number of contributing studies to assess for small differences among the approaches. Second, in the calculation of event rates using the incidence rate-based methods, we assumed complete follow-up of participants in each study. However, this method is robust to incomplete follow-up if the number of events and the amount of time contributed by each participant are known or it can be assumed that individuals lost to follow-up contribute no events or follow-up time and loss to follow-up is not differential between the treatment groups.

Conclusions

In this study, we demonstrated that choice of method among the ones presented here for continuous rate measures had little effect on inference. SMD, IRD, and IRR methods all gave qualitatively similar estimates of effect and suggest that the intervention was effective for both outcomes. However, choice of method clearly affected clinical interpretability. SMD, reportedly the standard method employed for analysis of rate measures of varying time duration, was not immediately interpretable. Stratified IRD allowed for clinical interpretability on an absolute scale. Stratified IRR or Poisson models allowed for clinical interpretability on a relative scale. For further discussion of the merits of absolute versus relative effects, we recommend that the reader consult additional references [10]. In addition as we have shown, failure to incorporate study indicators in the Poisson analysis may produce different (and inappropriate) estimates of treatment effect. (For an explanation of why we consider this an inappropriate approach, see the Appendix). We recommend that statistical software packages used for meta-analysis consider the addition of stratified IRD and IRR procedures.

Appendix

Table 5 demonstrates the need to perform analyses stratified by study when comparing event rates between treatments. A similar argument would apply to the comparison of risks. The principle demonstrated, among epidemiologists, would be called "confounding by study," and among statisticians might be more familiar as an example of "Simpson's Paradox." In brief, we have gen-

Table 5: Example of Confounding by Study.

	Control			Treated			
Study	Events	Person-time	Rate	Events	Person-time	Rate	Relative Rate
1	10	100	0.10	5	100	0.05	0.50
2	40	100	0.40	5	25	0.20	0.50
Total (ignoring "study")	50	200	0.25	10	125	0.08	0.32

erated a hypothetical example, in the table, of a situation in which the baseline (control) rates differ markedly between studies. In addition, the feature that generates the problem is that there is imbalance in the amount of person-time in the treatment and control groups in the second study, perhaps as a result of unequal allocation of subjects to the two conditions.

Within each study, the estimate of the relative risk is 0.5. Thus, any reasonable analysis that takes stratification by study into account (and averages the within-study treatment effects) would necessarily produce an average treatment effect of 0.5. Because of the associations noted above, the analysis ignoring study produces an estimated treatment effect of 0.32. This result clearly is not at all representative of the results within either of the individual studies. Note that this concept is *not* the same as the usual concept of "heterogeneity," which is generally used to refer to situations in which the treatment effect varies across studies. In our example, the treatment effect is constant across studies (on the relative rate scale), although the baseline rate varies dramatically between the studies.

Competing interests

None declared.

Authors' Contributions

JG conceived of the study, participated in the design and analysis of the study, wrote the manuscript. JB participated in the design of the study, performed the main statistical analysis, and participated in writing the manuscript. FW participated in the design and analysis of the study. All authors read and approved the final manuscript.

Acknowledgments

We would like to thank Doug Altman for his critical review of the manuscript. We would also like to acknowledge Russell Localio for sharing his STATA program on implementing stratified incidence rate differences for fixed- and random-effects models. This paper was presented at the XI Cochrane Colloquium, Barcelona, Spain, on October 31, 2003.

References

1. Egger Matthias, Smith George Davey, O'Rourke Keith: **Principles of and procedures for systematic reviews.** *Systematic reviews in health care: meta-analysis in context* 2nd edition. Edited by: Egger Matthias, Smith George Davey and Altman Douglas G. London, BMJ Publishing Group; 2001:23-42.
2. Thacker Stephen B: **Meta-analysis: a quantitative approach to research integration.** *JAMA* 1988, **259:**1685-1689.
3. Jadad Alejandro R, Moore R Andrew, Carroll Dawn, Jenkinson Crispin, Reynolds D John M, Gavaghan David J, McQuay Henry J: **Assessing the quality of reports of randomized clinical trials: is blinding necessary?** *Control Clin Trials* 1996, **17:**1-12.
4. Egger Matthias, Smith George Davey, Schneider Martin, Minder Cristoph: **Bias in meta-analysis detected by a simple, graphical test.** *BMJ* 1997, **315:**629-634.
5. Begg CB, Mazumdar M: **Operating characteristics of a rank correlation test for publication bias.** *Biometrics* 1994, **50:**1088-1099.
6. Victor S, Ryan SW: **Drugs for preventing migraine headaches in children (Cochrane Review). The Cochrane Library, Issue 4, 2003.** Chichester, UK, John Wiley & Sons; 2003.
7. Shea B, Wells G, Cranney A, Zytaruk N, Robinson V, Griffith L, Hamel C, Ortiz Z, Peterson J, Adachi J, Tugwell P, Guyatt G, The Osteoporosis Methodology Group, The Osteoposrosis Research Advisory Group: **Calcium supplementation on bone loss in postmenopausal women (Cochrane Review), Issue 4, 2003.** Chichester, UK, John Wiley & Sons; 2003.
8. Nannini L, Lasserson TJ, Poole P: **Combined corticosteroid and longacting beta-agonist in one inhaler for chronic obstructive pulmonary disease (Cochrane Review), Issue 4, 2003.** Chichester, UK, John Wiley & Sons; 2003.
9. Gray OM, McDonnell GV, Forbes RB: **Intravenous immunoglobulins for multiple sclerosis (Cochrane Review), Issue 4, 2003.** Chichester, UK, John Wiley & Sons; 2003.
10. Deeks Jonathan J, Altman Douglas G, Bradburn Michael J: **Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis.** *Systematic reviews in health care: meta-analysis in context* 2nd edition. Edited by: Egger Matthias, Smith George Davey and Altman Douglas G. London, BMJ Publishing Group; 2001:285-312.
11. Wolf Frederic M, Guevara James P, Grum Cyril M, Clark Noreen M, Cates Christopher J: **Educational interventions for asthma in children (cochrane review), in the Cochrane Library, Issue 1.** [http://www.update-software.com/abstracts/ab000326.htm].
12. DerSimonian R, Laird N: **Meta-analysis in clinical trials.** *Control Clin Trials* 1986, **7:**177-188.
13. Laird NM, Mosteller F: **Some statistical methods for combining experimental results.** *Int J Tech Assess in Health Care* 1990, **6:**5-30.
14. Rothman KJ, Greenland S: **Modern epidemiology.** 2nd edition. Philadelphia, Lippincott-Raven; 1998.
15. White H: **A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity.** *Econometrica* 1980, **48:**817-838.
16. Guevara James P, Wolf Frederic M, Grum Cyril M, Clark Noreen M: **Effects of educational interventions for self management of asthma in children and adolescents: systematic review and meta-analysis.** *BMJ* 2003, **326:**1308.

Pre-publication history

The pre-publication history for this paper can be accessed here:

http://www.biomedcentral.com/1471-2288/4/17/prepub