

General principles for systematic reviews and meta-analyses and a critique of a recent systematic review of long-acting β -agonists

Vernon M. Chinchilli, PhD *Hershey, Pa*

A *systematic review* attempts to summarize the scientific evidence related to treatment, causation, diagnosis, or prognosis of a specific disease. *Meta-analysis* refers to that portion of the systematic review that involves the statistical analysis. This perspective describes the appropriate steps in a systematic review and meta-analysis and critiques the recent systematic review on the effect of long-acting β -agonists on severe asthma exacerbations and asthma-related deaths. The authors of this systematic review identified 19 relevant studies and applied most of the methodological steps appropriately, although there is some concern about publication bias. The authors uncovered statistically significantly increased risk for long-acting β -agonists compared with placebo with respect to severe asthma exacerbations, life-threatening asthma exacerbations, and asthma-related deaths. Out of the 19 studies included in the systematic review, the Salmeterol Multicenter Asthma Research Trial provided 80% of the data and dominated the meta-analysis component, especially with respect to asthma-related deaths. (*J Allergy Clin Immunol* 2007;119:303-6.)

Key words: *Overviews, literature search, publication bias, funnel plots, fixed-effects models, random-effects models*

An *overview* (or *systematic review*) attempts to summarize the scientific evidence related to treatment, causation, diagnosis, or prognosis of a specific disease. An overview does not generate any new data; it reviews and summarizes already existing studies. *Meta-analysis* refers to that portion of the overview involving the statistical analysis of the selected studies. Overviews are important because there usually exist multiple studies that have addressed a specific research question, yet those studies

may differ with respect to design, patient population, quality, and/or results. Various guidelines for conducting and reporting on overviews in clinical trials and epidemiology are available.¹⁻³

Investigations into the safety issues of long-acting β -agonists, such as salmeterol, in treating asthma have raised serious concerns⁴ and have resulted in the issuance of a black box warning by the US Food and Drug Administration on salmeterol and salmeterol-fluticasone combination medications. This action has generated much discussion within the asthma research community.⁵⁻⁷ Recently, a systematic review with accompanying meta-analysis was published on the effect of long-acting β -agonists on severe asthma exacerbations and asthma-related deaths.^{8,9} The objective of this article is 2-fold: (1) describe the appropriate steps in performing an overview of clinical trials (Methods), and (2) critique the recent overview/meta-analysis on the effect of long-acting β -agonists on severe asthma exacerbations and asthma-related deaths (Results).

METHODS

Conducting an overview well requires a good deal of effort and care. There are 6 basic steps to an overview.

(1) Define a focused clinical question

If the clinical question is too broad, it may not be useful when applied to a particular patient. For example, "Is chemotherapy effective in cancer?" is too broad a question (the number of studies addressing this question could exceed 10,000). If the question is too narrow, there may not be enough evidence to answer the question. For example, "Is a particular asthma therapy effective in white females older than 65 years in central Pennsylvania?" is too narrow.

(2) Conduct a thorough literature search

The researcher should explore various sources for studies (throughout the world) that include bibliographic databases (MEDLINE, EMBASE, and so forth), conference proceedings, theses/dissertations, data banks of pharmaceutical firms, personal contacts, and unpublished reports. The pitfall that the researcher may encounter is publication bias. Studies with negative results (ie, the intervention is not found to be effective, or as effective as other treatments) sometimes are not published. In other words, an overview based only on published studies may be biased toward an overall

From the Department of Health Evaluation Sciences, Penn State College of Medicine.

Disclosure of potential conflict of interest: V. M. Chinchilli has patent licensing arrangements with the Childhood Asthma Research and Education Network of the National Heart, Lung, and Blood Institute in regard to a patent pending in the genetics of controller response.

Received for publication October 19, 2006; revised November 20, 2006; accepted for publication November 21, 2006.

Reprint requests: Vernon M. Chinchilli, PhD, Department of Health Evaluation Sciences, A210, Penn State College of Medicine, 600 Centerville Drive, Suite 2200, Hershey, PA 17033-0855. E-mail: vchinch@psu.edu. 0091-6749/\$32.00

© 2007 American Academy of Allergy, Asthma & Immunology
doi:10.1016/j.jaci.2006.11.632

positive effect. Publication bias sometimes is referred to as the “file-drawer problem” because studies with negative results tend to be filed in a drawer and not submitted for publication. Editors tend to prefer publishing positive studies in their journals, which contributes to publication bias. Most experts in the field consider publication bias to be the most difficult problem to overcome when conducting an overview.^{10,11}

Suppose there are some relevant studies with small sample sizes. If nearly all of them have a positive finding ($P < .05$), this may provide evidence of a publication bias because it is more difficult to show positive results with small sample sizes. Thus, there should be some negative results ($P > .05$) among the small studies. A funnel plot can be constructed to investigate this issue.^{10,11} The funnel plot consists of study sample size (vertical axis) versus P value or magnitude of effect (horizontal axis).

An excellent example of a thorough literature search is that conducted by a research group who developed an overview on studies investigating the dose-response effects of inhaled fluticasone propionate.¹² The researchers searched MEDLINE and EMBASE to identify 204 studies involving inhaled fluticasone propionate, and then they contacted the pharmaceutical manufacturer to determine if unpublished study reports were available.

(3) Apply inclusion/exclusion criteria

The researcher needs to establish eligibility criteria for the studies before conducting the meta-analysis. The researcher should base the inclusion/exclusion criteria on the design aspects of the trials, the patient populations, treatment modalities, and so forth that are congruent with the objectives of the overview.

Although somewhat subjective, researchers can grade the selected studies according to quality. One such example of a quality rating is the 5-point Jadad scale that ranges from 0 to 5¹³:

- Is the study described as randomized? (no, or yes but inappropriate method = 0 points; yes but no discussion of method = 1 point; yes and appropriate method = 2 points)
- Is the study described as double-blind? (no, or yes but inappropriate method = 0 points; yes but no discussion of method = 1 point; yes and appropriate method = 2 points)
- Is there a description of withdrawals/dropouts? (no = 0 points; yes = 1 point)

Based on the circumstances, a researcher may decide to exclude studies from the overview that do not reach a certain threshold on the quality scale. On the other hand, there are some concerns that scoring the quality of clinical trials for a meta-analysis can be very problematic.¹⁴

In the inhaled fluticasone propionate overview, the researchers specified the eligibility criteria to include only studies that were double-blind, randomized, placebo-controlled trials in patients with asthma at least 12 years of age, with more than 1 delivered dose of inhaled fluticasone propionate and with only 1 type of delivery system.¹² Of the 204 studies that the researchers initially identified, only 8 satisfied all of the eligibility criteria.

(4) Abstract/summarize the data

In most circumstances, the researcher easily can gather the relevant descriptive statistics (eg, means, SEs, sample sizes) from the reports on the eligible studies. Sometimes, older reports (say, before 1980) do not include variability estimates (eg, SEs). If possible, the researcher should attempt to contact the authors directly in such situations. This may not be successful, however, because some of the authors may no longer have the data or may no longer be alive. Ideally, the statistical analysis for a systematic review will be based on the raw data from each eligible study. This rarely occurs, however, because most authors are not willing to share their raw data or the raw data no longer are available.

(5) Perform a meta-analysis

The obvious advantage for performing a meta-analysis is that a large amount of data, pooled across multiple studies, can provide increased precision in addressing the research question. The disadvantage of a meta-analysis is that the studies can be very heterogeneous in their designs, quality, and patient populations; therefore, it may not be valid to pool them. Researchers invoke 2 basic statistical models for meta-analysis: fixed-effects models and random-effects models.^{15,16}

A fixed-effects model is more straightforward to apply, but its underlying assumptions are somewhat restrictive. It assumes that if all the involved studies had tremendously large sample sizes, then they all would yield the same result. In essence, a fixed-effects model assumes that there is no interstudy variability (study heterogeneity). The statistical model accounts only for intrastudy variability. A random-effects model, however, assumes that the eligible studies actually represent a random sample from a population of studies that address the research question. It accounts for intrastudy and interstudy variability. Thus, a random-effects model tends to yield a more conservative result—that is, wider CIs and less statistical significance—than a fixed-effects model.

A random-effects model is more appealing from a theoretical perspective, but it may not be necessary if there is very low study heterogeneity. A formal test of study heterogeneity is available. Its results, however, should not determine whether to apply a fixed-effects model or random-effects model. The test for study heterogeneity is very powerful and sensitive when the number of studies is large. It is very weak and insensitive if the number of studies is small. Graphical displays provide much better information about the nature of study heterogeneity. Some medical journals require that the authors provide the test of heterogeneity, along with a fixed-effects analysis and a random-effects analysis.

The basic step for a fixed-effects model involves the calculation of a weighted average of the treatment effect across all of the eligible studies. For a continuous outcome variable, the measured effect is expressed as the difference between sample treatment and control means. The weight is expressed as the inverse of the variance of the difference between the sample means. For a binary outcome variable, the measured effect usually is expressed as the logarithm of the estimated odds ratio. The weight is expressed as the inverse of the variance of the logarithm of the estimated odds ratio. The random-effects model for meta-analysis is much more complex than the fixed-effects model. A weighted analysis can be applied for the random-effects model, but the weights are calculated in a more complex manner to account for intrastudy and interstudy variability.

Graphical displays showing the estimated treatment difference and its CI for every study are very useful for evaluating treatment effects over time or with respect to other factors. Statistical diagnostics (sensitivity analyses) should be performed to investigate the validity and robustness of the meta-analysis. Suppose there are K studies that meet the eligibility criteria for inclusion in the meta-analysis. This is performed by applying the meta-analytic approach to subsets of the K studies, and/or applying the leave-1-out method. The steps for the leave-1-out method are as follows:

- Remove the first of the K studies and conduct the meta-analysis on the remaining $K - 1$ studies
- Remove the second of the K studies and conduct the meta-analysis on the remaining $K - 1$ studies
- Continue this process until there are K distinct meta-analyses (each with $K - 1$ studies)

If the results of the K meta-analyses in the leave-1-out method are consistent, then there is confidence that the overall meta-analysis is robust. The likelihood of consistency increases as K increases. Rather

than invoke the leave-1-out method, researchers often prefer to perform a sensitivity analysis by applying the meta-analysis to subsets of studies on the basis of high-quality versus low-quality studies, randomized versus nonrandomized studies, early studies versus late studies, and so forth.

(6) Disseminate the results

Often, researchers overlook the importance of disseminating the results and how to report the results of an overview. Some articles have appeared in the literature that provide guidelines on how to report the results.¹⁻³ Cochrane Reviews (<http://www.cochrane.org>) have become very popular and are based on the best available information about healthcare interventions. They explore the evidence for and against the effectiveness and appropriateness of treatments (medications, surgery, education, and so forth) in specific circumstances.

RESULTS

The authors of the recent systematic review and meta-analysis on the effect of long-acting β -agonists on severe asthma exacerbations and asthma-related deaths pursued the 6 steps described in the Methods section for this communication.⁸ The researchers searched MEDLINE, EMBASE, CINAHL, and Cochrane databases for studies published between 1966 and 2005 on long-acting β -agonists. The researchers also investigated the references of the identified publications and explored the US Food and Drug Administration Web site. They did not, however, describe any attempts to scan abstracts from meetings of important professional societies, such as the American Academy of Allergy, Asthma & Immunology, the American Thoracic Society, and so forth. It is possible that some studies were presented at such meetings, but these studies never made it into print. Another possible source of studies that the authors could have explored, but did not, was to contact the pharmaceutical manufacturers (similar to what the researchers did for the overview of a dose-response effect for inhaled fluticasone propionate¹²). The authors admit that one of the limitations of their overview is that "Our analysis was based mainly on published literature and therefore may be subject to publication bias."

With respect to eligibility criteria, the researchers required that studies be randomized, controlled trials of long-acting β -agonists compared with placebo and with at least 3 months of patient follow-up. In addition, the studies needed to include 1 or more of the following outcomes:

- Severe asthma exacerbations requiring hospitalization
- Life-threatening asthma exacerbations requiring intubation and ventilation
- Asthma-related deaths

Out of the approximately 5000 studies initially identified through the search, only 19 studies met all of these criteria. The researchers rated the quality of each of the 19 studies on a 10-point scale (0 through 9) based on

- Adequate randomization procedures that were nondiscoverable

- Double-blinding
- Reporting of dropouts/withdrawals and adherence to the intention-to-treat principle

Each of the 19 studies received a quality score of 8 or 9, so the researchers did not perform any sensitivity analyses based on quality ratings.

A total of 33,826 participants, with 16,848 patient-years of follow-up, were accumulated across these 19 studies. The SMART trial dominated in that it involved 26,353 participants (80% of the total in the overview).⁴ The long-acting β -agonist and placebo groups did not differ significantly with respect to age, sex, and dropout rates, and approximately 15% of the participants were African American. The researchers applied a fixed-effects model for the meta-analysis of estimating the odds ratio and the risk difference for each of the 3 outcomes described. The researchers did not apply any random-effects models for the meta-analysis, but that may not have been necessary because there appeared to be very low levels of study heterogeneity ($P > .85$ for severe asthma exacerbations requiring hospitalization and for life-threatening asthma exacerbations; P value on heterogeneity not reported for asthma-related deaths).

With respect to severe asthma exacerbations requiring hospitalizations, 12 of the 19 studies provided data for the meta-analysis (the SMART study was one of those that did not). The pooled estimate of the odds ratio was 2.6 (95% CI, 1.6–4.43) and the pooled estimate of the risk difference was 0.7% (95% CI, 0.1% to 1.3%), indicating that the long-acting β -agonist group had significantly higher risk. Only 1 study on its own, however, displayed a statistically significant higher risk for the long-acting β -agonist group.

With respect to life-threatening asthma exacerbations, 7 of the 19 studies provided data for the meta-analysis (including the SMART study). The pooled estimate of the odds ratio was 1.8 (95% CI, 1.1–2.9) and the pooled estimate of the risk difference was 0.12% (95% CI, 0.01% to 0.3%), indicating that the long-acting β -agonist group had significantly higher risk. Not one study on its own (including the SMART study), however, displayed a statistically significant higher risk for the long-acting β -agonist group.

With respect to asthma-related deaths, 14 of the 19 studies provided data for the meta-analysis (including the SMART study). The SMART study dominated this meta-analysis because there were very few asthma-related deaths uncovered in the other 13 studies. In the SMART study, there were 13 asthma-related deaths among 13,174 participants in the long-acting β -agonist group and 3 asthma-related deaths among 13,179 participants in the placebo group. The authors only provided the meta-analytic estimate and CI for the risk difference, which was 0.07% over a 6-month period (95% CI, 0.01% to 0.1%), indicating that the long-acting β -agonist group had significantly higher risk. Given the domination by the SMART trial, however, it may be incorrect to claim that this constitutes a systematic overview to investigate the risk of long-acting β -agonists on asthma-related deaths.

Except for the possibility of a publication bias, the researchers conducted the overview and meta-analysis relatively well. They made 1 particular assertion, however, in their Discussion section that is unfounded:

“In this meta-analysis, the risks for severe exacerbations and asthma-related deaths were increased by 2- to 4-fold. However, we must also look at the absolute risk increase to put this into clinical perspective. We found that the absolute increase in asthma-related deaths was estimated to be 0.06% to 0.07% over six months, indicating that long-acting β -agonists cause an excess of approximately 1 death per 1000 patient-years of use. Salmeterol is one of the most widely prescribed medications in the world, with an estimated 3.5 million adults treated in the United States in 2004. This indicates that salmeterol may be responsible for approximately 4000 of the 5000 asthma-related deaths that occur in the United States each year.”

The major problem with respect to this assertion is the assumption that an added risk of 0.06% to 0.07% over 6 months is comparable to 1 death per 1000 patient-years. Because this assumption is based on data from 6 months of follow-up and very small numbers of observed deaths, it is unknown whether the risk remains constant beyond 6 months, which in turn means that 1 death per 1000 patient-years may be quite inaccurate. Also, the assumption relies on the point estimate of 0.07% for the risk difference and ignores the 95% CI, which indicates that the actual risk difference may be as low as 0.01%.

REFERENCES

1. Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized controlled trials in health care from the PotSDam consultation on meta-analysis. *J Clin Epidemiol* 1995;48:167-71.
2. Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF, for the QUORUM Group. Improving the quality of reports of meta-analyses of randomized controlled trials. *Lancet* 1999;354:1896-900.
3. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al, for the Meta-analysis of Observational Studies in Epidemiology (MOOSE) Group. Meta-analysis of observational studies in epidemiology. *JAMA* 2000;283:2008-12.
4. Nelson HS, Weiss ST, Bleecker ER, Yancey SW, Dorinsky PM. The Salmeterol Multicenter Asthma Research Trial: a comparison of usual pharmacotherapy for asthma or usual pharmacotherapy plus salmeterol. *Chest* 2006;129:15-26.
5. Nelson HS. Is there a problem with inhaled long-acting β -adrenergic agonists? *J Allergy Clin Immunol* 2006;117:3-16.
6. Szeffler SJ, Whelan GJ, Leung DYM. “Black box” warning: wake-up call or overreaction? *J Allergy Clin Immunol* 2006;117:26-9.
7. Pearlman DS. *Primum non nocere* and “black box” warnings. *J Allergy Clin Immunol* 2006;117:30-1.
8. Salpeter SR, Buckley NS, Ormiston TM, Salpeter EE. Meta-analysis: effect of long-acting β -agonists on severe asthma exacerbations and asthma-related deaths. *Ann Intern Med* 2006;144:904-12.
9. Glassroth J. The role of long-acting β -agonists in the management of asthma: analysis, meta-analysis, and more analysis. *Ann Intern Med* 2006;144:936-7.
10. Copas J, Shi JQ. Meta-analysis, funnel plots and sensitivity analysis. *Biostatistics* 2000;1:247-62.
11. Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Comparison of two methods to detect publication bias in meta-analysis. *JAMA* 2006; 295:676-80.
12. Holt S, Suder A, Weatherall M, Cheng S, Shirtcliffe E, Beasley R. Dose-response relation of inhaled fluticasone propionate in adolescents and adults with asthma: meta-analysis. *BMJ* 2001;323:1-8.
13. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJM, Gava-ghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996;17:1-12.
14. Jüni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999;282:1054-60.
15. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177-88.
16. Schmid JE, Koch GG, LaVange LM. An overview of statistical issues and methods of meta-analysis. *J Biopharm Stat* 1991;1:103-20.