

Meta-analysis in Cancer Epidemiology

Robert D. Morris

Division of Epidemiology, Medical College of Wisconsin, Milwaukee, Wisconsin

Meta-analysis has seen increasing use as a tool in epidemiology over the past five years. Although this method is relatively well accepted for use in clinical trials, its use has proved somewhat more controversial in epidemiology. If meta-analysis is viewed as an evolutionary improvement over the review article, it may become more widely acceptable. Meta-analyses should incorporate the concern for study quality and differences in study design seen in classic review articles with the concern for rigor, objectivity, and quantitative precision characteristic of meta-analysis. Available tools for consideration of differences among studies are described with several examples from the literature. The extent to which various methods are used in published meta-analyses is described. Methods for assessing publication bias, and tools for combining dose-response data, are discussed also. Evaluation of risk factors and protective factors for cancer must be based on the weight of the evidence. Tools such as meta-analysis are essential if we are to interpret the vast number of completed studies in cancer epidemiology. — *Environ Health Perspect* 102(Suppl 8):61–66 (1994)

Key words: Meta-analysis, epidemiology methodology

Introduction

Improving our understanding of cancer epidemiology requires new, more powerful tools for evaluating the health effects of exposures to possible carcinogens and anti-carcinogens. Cancer epidemiologists often encounter difficulty in obtaining samples of sufficient size to detect the effect of exposures. Noise in the data often obscures the effect of those exposures. Meta-analysis represents an important innovation for increasing statistical power in epidemiologic studies.

Some epidemiologists seem to view meta-analysts as a form of intellectual parasite who collects the hard work of others, extracts a few numbers from each study, performs some inappropriate calculations and takes credit for the conclusions. Wherever one places meta-analysts within the epidemiologic food web, they require a rich supply of completed studies for sustenance. In cancer epidemiology, the resources for meta-analysis have been abundant. Table 1 lists the annual publication rate of epidemiologic studies listed in MEDLINE for various types of neoplasms. This table demonstrates a steady, almost monotonic increase in the rate of publication in every area of cancer epidemiology over the past 27 years.

Although meta-analysts are opportunistic feeders, they have been somewhat slow to take advantage of the niche available in

cancer epidemiology. MEDLINE lists only 32 meta-analyses in cancer epidemiology, 28 of which were published in the past three years.

The slow introduction and use of meta-analysis in epidemiology in general and cancer epidemiology in particular reflects the difficulties associated with meta-analysis in this context. The most extensive use of meta-analysis in medical research has involved clinical trials. The relative consistency of study designs and similarity of outcome measures has facilitated meta-analysis in this area. Epidemiologic studies, on the other hand, use a wide range of study populations and methods with a variety of measures of exposure and outcome. Consequently, they are more difficult to combine and meta-analysis is less well accepted in epidemiology than in clinical trials.

The Problem of Evaluating the Weight of the Evidence

Evaluating a set of studies in cancer epidemiology that focuses on a specific risk factor requires some method for combining the results of these studies. Any method must follow the general model described by Equation 1, a simple, weighted average of results.

Table 1. Average number of studies per year in cancer epidemiology.

Site	Year of publication			
	1966–75	1976–82	1983–87	1988–92
Bladder	7	5	12	19
Colon	9	22	32	24
Breast	26	47	51	75
Prostate	3	4	5	8
Brain	11	9	11	15
Lung	28	27	47	55
Liver	9	14	17	18
Uterus	12	16	20	19
Total	109	152	201	245

$$\bar{y}_w = \sum w_i y_i / \sum w_i$$

where y is the summary statistic from study i , w_i is the weight applied to study i , and \bar{y}_w is the pooled summary statistic.

The specific method used to determine the weights and the method used to determine the level of confidence associated with the pooled relative risk estimate differentiates the methods used in combining study results. These methods fall into two broad categories, qualitative and quantitative.

The classic review article takes the qualitative approach in determining what weights to apply when combining study results. Review articles generally do not employ a strict protocol for identification of all completed studies. Studies that are not included are given an implicit weight of zero. In evaluating studies that are identified, the approach taken in most review articles involves excluding or discounting studies from consideration based upon the perceived potential for bias and confounding. Those studies felt to be fatally flawed are given an implied weight of zero. The remaining studies are combined according to a vague, implicit weighting scheme that tends to give greater weight to large prospective studies based upon the assumption of greater potential for bias in retrospective studies. Review articles usually conclude with an assertion as to the presence or absence of a true risk rather than a specific estimate of the relative risk of cancer. The emphasis in review articles is on the validity of individual studies rather than the accuracy of the final risk estimate.

A simple quantitative approach to pooling study results emphasizes precision in combining study results and relies on an

This paper was presented at the 4th Japan-US Biostatistics Conference on the Study of Human Cancer held 9–11 November 1992 in Tokyo, Japan.

Address correspondence to Robert D. Morris at Division of Epidemiology, Medical College of Wisconsin, 8701 Watertown Plank Road, Milwaukee, WI 53226. Telephone (414) 456-8200. Fax (424) 257-8353. E-mail Internet rdmorris@post.its.mcw.edu

assumption that the group of studies being pooled represents a homogeneous set of experiments. The simplest weighting scheme gives a uniform weight of one to every study. Uniform weighting reduces Equation 1 to a simple, unweighted average. This is clearly inappropriate and uninformative. If the studies to be combined are truly homogeneous, the best linear, unbiased estimator of the true mean is derived using weights based upon precision. Precision is expressed as the inverse of the variance for the summary statistic. This model is often referred to as the Peto method for meta-analysis (1).

The homogeneity assumption in meta-analysis gives rise to the most important criticism of meta-analysis. Critics charge that a meta-analysis is simply precision masquerading as validity. In other words, meta-analysis produces a single estimate of relative risk with confidence intervals that are substantially narrower than those in any individual study. This suggests a great improvement in precision, but may fail to consider the importance of validity in evaluating a group of studies. Some degree of heterogeneity is inevitable when different investigators use different methods and study populations, despite the fact that they may be exploring the same possible causal relationship.

The classic review paper takes the opposite approach, focusing on validity to the exclusion of precision. The presumption that weaknesses or potential weaknesses in the design of individual studies should be the overriding concern in combining study results drives the conventional review. In drawing a conclusion based on a qualitative assessment of individual studies, the author of a review article conducts a crude form of meta-analysis with a simplistic, unspecified weighting protocol. As a consequence, the reviewer sacrifices precision and scientific rigor over a concern about the heterogeneity among studies.

Clearly, both of these extreme approaches are flawed. A qualitative review article is prone to bias and tends to be inefficient in its use of the available data. A meta-analysis that focuses on a single, numerical result without careful consideration of the issues of validity and the factors that differentiate the individual studies is inadequate and open to misinterpretation. The ideal meta-analysis combines the rigor of the purely quantitative meta-analysis with the concern for variations in study design exhibited in review articles.

Meta-analysis and Its Application in Cancer Epidemiology

Any meta-analysis includes four components: identification of a set of combinable studies, extraction of summary data, pooling of summary data, and delineation of differences in study design with an evaluation of the impact of those differences. Let us consider each of these, paying careful attention to the concerns raised above.

Study Selection and Publication Bias

The first step in any meta-analysis is the identification of an appropriate set of studies for pooling. Although computerized literature searches have greatly facilitated this process, they inevitably miss relevant published studies. Inconsistencies in the use of key words require broad, inclusive searches with careful review of identified studies to locate additional studies. The reference list from each study should be checked to identify as many of the pertinent published studies as possible. A clearly defined protocol should then be used to identify studies appropriate for meta-analysis.

The problem of unpublished studies creates a potential for bias with no simple solution. If a decision on the part of a researcher to submit a study for publication or a decision on the part of a journal to accept a publication is influenced by the direction of the study results, a bias is introduced. This systematic failure by investigators or journals to publish studies is referred to as "publication bias." Although publication bias has been amply demon-

strated for clinical trials (2), it has not been systematically investigated for epidemiologic studies. The potential for bias in cancer epidemiology is arguably high. Case-control studies may investigate a wide range of hypotheses. The lack of a strong incentive to publish negative results from these studies may lead to publication bias.

One approach that has been advocated to minimize bias is the registration of ongoing epidemiologic studies (3). A comprehensive registry of epidemiologic studies should help to minimize publication bias. A registry of studies in cancer epidemiology has been initiated by the International Agency for Research on Cancer (DM Perkins, personal communication). The success of this registry will help determine if this approach is feasible.

In the absence of a registry of epidemiologic studies, it is difficult, if not impossible, to quantify publication bias. It is possible, however, to get some insight as to the presence or absence of publication bias. One can reasonably assume that larger and consequently more expensive studies are less likely to go unpublished than smaller, less precise studies. A plot of published study results as a function of some measure of the precision of the results (e.g., standard error of the log relative risk) should yield a funnel-shaped scatter plot with a decrease in the scatter of results as precision increases (4). A truncation of the lower half of the funnel would imply that small negative studies had not been published.

Figure 1 shows an example of a funnel plot constructed using data from a study of colorectal cancer and alcohol consumption

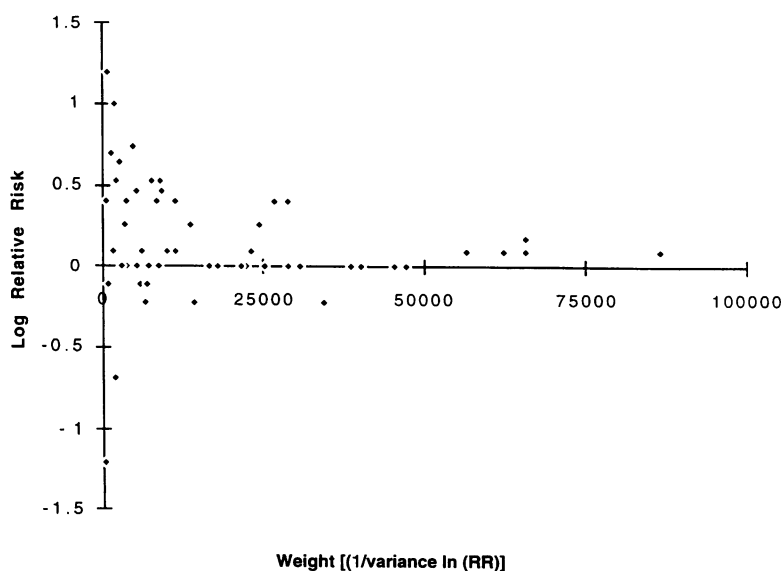


Figure 1. A funnel plot for epidemiologic studies of alcohol consumption and colorectal cancer.

(5). This plot is somewhat truncated, suggesting a degree of publication bias. In fact, if we divide these studies according to tertile of precision, the relative risk drops from 1.35 to 1.15 to 1.08 with increasing precision. This suggests that some publication bias may exist for the smaller, less precise studies. Begg and Mazumdar (6) have recently proposed a method for detection of publication bias using an adjusted rank correlation test that represents a statistical analog of the funnel plot.

One can also estimate the possible impact of publication bias on a specific meta-analysis. Calculation of the size of a negative study that would be required to increase the probability of a Type I error to a level considered nonsignificant will provide an estimate of the extent of publication bias necessary to invalidate the conclusions of a meta-analysis. Sugita et al. (7) have proposed a method that infers the magnitude of missing studies based upon an assumed distribution of study results and adjusts the meta-analysis accordingly.

Data Extraction

A meta-analysis requires the abstraction of comparable measures of risk along with the variance associated with the summary statistic for risk from each study. Again, the variability among studies poses a problem for the meta-analyst. Risk estimates may come in many forms, including odds ratios and relative risks from binary exposure data or logistic regression coefficients from continuous exposure measures. Outcomes may include mortality, a wide range of indicators of morbidity, or biomarkers for subclinical disease.

Perhaps the most difficult task of an epidemiologic meta-analysis is to determine comparable measures of exposure among studies. The meta-analyst must make decision based upon knowledge of the relevant epidemiology to determine which summary statistics are comparable.

Determination of weights for meta-analysis depends upon extraction of variance estimates. Many studies do not provide variance estimates and some do not provide sufficient statistics for calculating variance estimates. Calculation of variances from confidence intervals or *p* values is relatively straightforward. Often, however, particularly in older studies, one must reconstruct the original data to calculate variances. Occasionally, even this is not possible, and the original data must be sought from the investigator.

Some investigators choose to pool data rather than summary statistics and analyze

the results as if they were the product of one large study. This approach is inappropriate unless the authors specifically account for the effects of the differences among studies in their analysis. One could argue that this is not, in fact, a true meta-analysis, since a meta-analysis, by definition, involves combining the results of multiple studies.

Many judgments are made in the selection and rejection of papers, the evaluation of study quality, and the extraction of data in the process of meta-analysis. Persons extracting data from individual studies must follow a strict protocol that is clearly defined in advance. Observer errors, occasionally the result of bias, can be reduced by arranging for all decisions and data extraction to be done by two independent investigators with differences settled by conference and, if necessary, by a third party.

Pooling of Study Results

Although the term meta-analysis often is used to refer only to the computational synthesis of results, these calculations tend to be the most straightforward and least time consuming aspect of the entire procedure. Nonetheless, the specific method chosen may affect a study's conclusions (8). Several methods are available for pooling results. Each method calculates a weighted average of the relative risk estimates from the original studies. The assumptions underlying the meta-analysis determine the weights to be used in the calculation.

These assumptions relate, in part, to the degree of heterogeneity among studies and the sources of that heterogeneity. Study results differ as a consequence of within-study variation and among-study variation. Within-study variation reflects variability in response for subjects investigated according to a single study design. The homogeneity assumption implies that variation among subjects is the primary source of variation. Among-study variation arises from differences in protocol among studies. These interstudy differences represent the consequences of fixed and random effects on subject response.

The Peto method (1) requires an assumption of homogeneity. Although it provides a test for the validity of this assumption, it does not provide any adjustment to account for that heterogeneity. This is the most widely used method for pooling studies in cancer epidemiology. Although the assumption of homogeneity is convenient, it is not entirely valid.

Use of the random effects model in pooling study results provides a purely

quantitative method of allowing for heterogeneity. The best known implementation of the random effects model for meta-analysis was developed by Dersimonian and Laird (8). The Dersimonian and Laird (D&L) method uses Cochran's *Q* and delta statistics to evaluate and adjust for heterogeneity in the variance among studies and is therefore referred to occasionally as the Cochran, Dersimonian, and Laird method. In calculating the weights for Equation 1, the random effects model adds a random component to the variance estimates before taking their inverse.

$$Q = \sum w_i (y_i - \bar{y}_w)^2, \tag{2}$$

where *Q* is Cochran's statistic for heterogeneity and

$$\bar{y}_w = \sum w_i y_i / \sum w_i \tag{3}$$

is the pooled estimate based on the assumption of heterogeneity with weights given by

$$w_i = (S_i^2)^{-1}, \tag{4}$$

where *S_i²* is the variance of the relative risk estimate for study *i*.

The expected value of *Q* is given by the equation

$$E(Q) = (k-1) + \sigma_A^2 (\sum w_i - \sum w_i^2 / \sum w_i), \tag{5}$$

where the random component of variance σ_A^2 is defined as

$$\max \left\{ 0, \frac{Q - (k-1)}{(\sum w_i - \sum w_i^2 / \sum w_i)} \right\} \tag{6}$$

The net effect of the D&L model is to decrease the weight on the most precise estimates of relative risk and to increase the variance associated with the pooled estimate of relative risk. Although this method addresses the question of heterogeneity mathematically, it does not deal directly with the problem of identifying the source of variation among studies. A meta-analysis that simply presents the pooled result of a group of studies without considering the impact of differences in study design on relative risk estimates represents a misuse of this method.

Many epidemiologic studies consider exposure in the form of a dichotomous variable. The summary result for these studies is an estimate of the relative risk associated with exposure. Often, however, the researchers will evaluate a dose response

relationship. Pooling dose response data will help determine if a causal link exists between the exposure and cancer. Greenland and Longnecker (9) describe a method for pooling the coefficients from dose-response relationships. This is developed further by Berlin, Longnecker, and Greenland (unpublished data).

Meta-analysis, in addition to improving our ability to identify small but significant relative risks in the context of inconsistent research findings, enhances our capacity to interpret negative results. When a meta-analysis yields nonsignificant results, an analysis of the potential for type II error associated with these results can provide useful insight into the need for additional studies.

Table 2 lists the results from a recent meta-analysis (9) of seven case-control and cohort studies that investigated the association between exposure to chlorination by-products in drinking water and cancer. The right side of the table lists the results of power calculations for the nonsignificant studies. These results help determine if the lack of significance associated with one cancer site represents a true lack of association, or a lack of statistical power. For example, these results suggest that exposure to chlorination by-products does not substantially increase the risk of lung cancer; but the available data clearly do not provide adequate power to reject the hypothesis of a small increase in the risk for brain cancer as a consequence of exposure to chlorination by-products.

Evaluation of Bias and Confounding

Heterogeneity among study results arises as a consequence of differences in study design that have not been accounted for. Confounding, interaction, and bias account for differences in results among studies, which are not due to chance. (If we accept that biologic mechanisms are deterministic, all differences among study results reflect a failure to account fully for the entire spectrum of causal factors.) The thorough, well-conceived meta-analysis will attempt to use available statistical tools to evaluate and, to the extent possible, quantify those factors that contribute to heterogeneity among studies. Several approaches are available.

Under ideal conditions, the impact of specific biases and confounders would be known with some precision and could be quantified. This information would allow for the quantitative adjustment of risk estimates before pooling. Unfortunately, the data needed to determine the correct

Table 2. Site-specific meta-analysis with power calculations for the association of chlorinated drinking water with cancer.^a

Site	n ^c	Relative risk estimate	95% confidence limits		p	Power ^b for detection of specified relative risks (α=0.05)		
						1.20	1.40	1.60
Bladder	7	1.21	1.09	1.34	<0.0001	— ^d	—	—
Brain	2	1.29	0.53	3.14	0.56	0.02	0.04	0.07
Breast	4	1.18	0.90	1.54	0.24	0.04	0.25	0.62
Colon	7	1.11	0.91	1.35	0.32	0.18	0.34	0.75
Colorectal	8	1.15	0.97	1.37	0.10	0.32	0.61	0.96
Esophagus	5	1.11	0.85	1.45	0.43	0.08	0.39	0.76
Kidney	4	1.16	0.89	1.51	0.23	0.05	0.30	0.67
Liver	4	1.15	0.94	1.40	0.16	0.06	0.50	0.91
Lung CA	5	1.01	0.86	1.18	0.94	0.57	0.98	>0.99
Pancreas	6	1.05	0.91	1.22	0.48	0.43	0.97	>0.99
Rectum	6	1.38	1.01	1.87	0.04	—	—	—
Stomach	6	1.14	0.94	1.38	0.19	0.08	0.56	0.93

^aFrom Morris et al. (9). ^bPower is defined as 1-β where β is the probability of Type II error at the given levels of true relative risk (RR) and significance. ^cNumber of studies evaluating specific cancer site. ^dStatistical power was not calculated if the meta-analysis was significant.

adjustment are the same data included in the specific meta-analysis. Consequently, quantitative adjustment for confounding and bias creates an impression of precision that is not entirely valid.

The meta-analyst can draw on several tools in an effort to evaluate the factors that contribute to heterogeneity in study results. These tools are similar to those used to evaluate confounding, interaction, and bias within the context of a single experiment and include stratification, meta-regression, sensitivity analysis, and quality scoring.

Stratified analysis is the most common tool in meta-analysis for evaluating study heterogeneity. With studies grouped based upon differences in study design, the investigator can pool the groups separately to determine if these differences (e.g., adjustment for a specific confounder) have an impact on relative risk estimates. Table 3 shows an extensive stratified analysis from a meta-analysis of oral contraceptives and breast cancer (10).

If the studies differ according to a continuous or ordinal independent variable, meta-regression may be appropriate. As described by Greenland (11), this method allows for evaluating the association between study results and a covariate using weighted least squares regression techniques based on the general relationship:

$$\ln(R) = B_0 + B_1 M_1 + e, \quad [7]$$

where R is the relative risk, B_0 is the baseline risk, M_1 is the effect modifier with coefficient B_1 , and e is an error term. Both the effect and effect modifiers are weighted by the reciprocal of the variance for each

study. This can be readily extended to a multivariate model.

Meta-regression could be used, for example, in the meta-analysis of a group of studies investigating associations between air pollution and lung cancer. To test the hypothesis that relative risk estimates might be influenced by the altitude of the city in which the study was conducted, the meta-analyst could include altitude in a meta-regression of these results. As in single experiments, regression is a more powerful tool than simple stratification. Unfortunately, use of meta-regression often is not feasible and the meta-analyst must resort to stratification.

Sensitivity analysis, a third tool for evaluating the impact of study heterogeneity on study results, has seen limited use in meta-analysis. If the presence of bias or confounding is suspected in a study or group of similar studies, the analyst can simulate the impact of a various levels of bias or confounding on study results. By simulating the relationship between confounding or bias and study results, the appropriateness of ignoring the effect of concern can be evaluated. Table 4 shows the results of a sensitivity analysis based upon various degrees of assumed confounding from a study that evaluated the association of smoking with cervical cancer (12).

Ultimately, there is likely to be a variety of factors affecting study results that are extremely difficult to quantify. Nonetheless, experience tells us that some study designs are more prone to bias than others. Authors of conventional review articles use implicit subjective criteria to evaluate differences in study design and determine the

Table 3. Stratification analysis from a meta-analysis of breast cancer and oral contraceptives.^a Summary relative risk (RR) estimates among case-control and follow-up studies for different characteristics of oral contraceptive (OC) use.

Characteristics of OC use	Random effect model Summary estimates RR (95% CI)	Test for heterogeneity	
		χ^2	<i>p</i>
Case-control studies			
Any OC use			
All studies	1.06 (0.98–1.14)	50.62	0.005
Cases accrued after 1975	1.08 (0.99–1.18)	44.71	0.005
Cases accrued after 1980	1.06 (0.93–1.21)	23.52	0.009
Premenopausal cases only ^b	1.17 (0.95–1.45)	13.77	NS
Duration of OC use			
< 1 year	0.97 (0.83–1.14)	20.24	0.027
≥ 10 years	1.14 (0.90–1.42)	15.70	NS
≥ 10 years (cases accrued after 1980)	1.22 (0.91–1.63)	13.30	0.037
Duration of OC use before first full-term pregnancy			
≤ 1 year	1.07 (0.78–1.47)	5.55	NS
≥ 4 years in women < 45 years old	1.72 (1.36–2.19)	11.69	NS
≥ 4 years in women < 45 years old (cases accrued after 1980)	1.73 (1.29–2.33)	8.86	NS
Time since first OC use			
≥ 12 years	1.08 (0.95–1.22)	22.71	0.01
Family history of breast cancer			
Yes	1.14 (0.67–1.96)	2.77	NS
No	1.44 (1.09–1.88)	3.49	NS
Follow-up studies			
Any OC use	1.06 (0.92–1.22)	6.51	NS

NS, Not significant. ^aFrom Romieu et al. (10). ^bFor the purposes of this analysis, premenopausal was defined as age ≤ 45 years.

unstated weighting criteria for epidemiologic studies. The goal of quality scoring in meta-analysis is to develop an explicit, objective scoring criterion to allow for quantification of the major components of study quality.

Quality scores provide a ranking system for the studies based upon methodology and the potential for bias and confounding. Although quality scoring has been used extensively in meta-analyses of clinical trials, no well-defined protocol exists for quality scoring of epidemiologic studies. A variety of quality scoring systems has been developed for epidemiologic meta-analysis (5,13,14). Two studies have demonstrated associations between relative risks and quality scores (1,14).

Summary and Conclusions

In the past 26 years, more than 4000 studies have been published in cancer epidemiology. Identification of cancer risk factors depends upon our ability to distill the results of these studies in a meaningful way. The traditional method for evaluating a group of studies, the review article, does not aspire to the same standards of scientific objectivity seen in the original epidemiologic studies. In some respects, a review article is simply a crude meta-analysis

that does not specify or follow a rigorous protocol in selecting or combining study results.

In using a rigorous protocol, a meta-analysis reduces the potential for bias present in review articles. Quantitative procedures for combining results also minimize the potential for bias while making the maximum use of information in the original studies. Although a meta-analysis can estimate a relative risk with far more precision than the individual studies it contains, that precision should not be mistaken for validity. The challenge for meta analysts is to fully consider differences in study design while taking advantage of the positive attributes of meta-analysis.

Qualitative review articles also may include evidence from outside of the epidemiologic literature. Extending meta-analysis to fully consider the weight of the evidence demands some effort to include the findings from nonepidemiologic literature, particularly toxicologic studies, as appropriate. This information can assist in evaluating the likelihood of a causal association based upon the issue of biological plausibility.

Table 5 lists the methods used in 19 published meta-analyses of cancer epidemiology. Of these, the majority used pooling

Table 4. Sensitivity analysis from a meta-analysis of cervical cancer and cigarette smoking.^a Summary of the results of graphical and variance-based meta-analyses and sensitivity analysis.^b

Sensitivity analysis	Graphical meta-analyses OR (95% CI)	Variance-based meta-analyses OR (95% CI)
No confounding		
Never smoked	1.00	1.00
Former smoker	1.18 (0.98–1.42)	0.98 (0.80–1.21)
Current smoker	1.81 (1.54–2.12)	1.69 (1.50–1.91)
Low degree of confounding (10%)		
Never smoked	1.00	1.00
Former smoker	1.06 (0.90–1.25)	0.88 (0.73–1.06)
Current smoker	1.63 (1.41–1.88)	1.52 (1.36–1.69)
High degree of confounding (30%)		
Never smoked	1.00	1.00
Former smoker	0.83 (0.73–0.95)	0.69 (0.60–0.80)
Current smoker	1.27 (1.14–1.42)	1.18 (1.09–1.28)

^aFrom Licciardone et al (12). ^bSensitivity analysis was used to address the degree to which the unadjusted ORs may have overestimated the true ORs because of unmeasured confounding variables.

of data to combine results. Only four used the D&L method. Stratification analysis was commonly used to evaluate the impact of differences in study design on results. Few of the studies evaluated the impact of potential biases and only four used quality scoring. Adoption of newer methods in meta-analysis together with more widespread use of existing methods should yield higher quality meta-analyses.

Meta-analysis provides an evolutionary advance over the conventional review article. Ongoing efforts to improve meta-analysis will be required if it is to become widely used and accepted as a tool in cancer epidemiology. Many of the existing problems in meta-analysis involve the inadequate capability to consider differences in study design and study quality. A meta-analysis must take into consideration the strengths and weaknesses of the literature while employing an explicit, objective, and quantitative methodology for pooling these

Table 5. Methods used in 19 published meta-analyses in cancer epidemiology.

Combining results	
Data pooling	8
Peto method	7
Dersimonian and Laird	5
Dose-response	4
Evaluation of confounding	
Stratification	14
Sensitivity	1
Evaluation of bias	
Quality scoring	4
Stratification	2
Sensitivity	3

studies. This approach to combining studies helps to minimize the impact of the biases of a particular reviewer on the interpretation of available studies. Overall, we must develop approaches to considering the weight of the evidence that employ the most rigorous methods available for minimizing bias while making the most efficient use of all available data.

REFERENCES

1. Yusuf S, Peto R, Lewis J, Collins R, Sleight P. Beta blockade during and after myocardial infarction: an overview of the randomized trials. *Prog in Cardiovasc Dis* 27:335-371 (1985).
2. Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results: follow-up of applications submitted to two institutional review boards. *JAMA* 267:374-378 (1992).
3. National Research Council. *Environmental Epidemiology: Data Gaps, Resource Needs and Research Opportunities*. National Academy of Sciences, in press.
4. Light RJ, Pillemer DB. *Summing Up: The Science of Reviewing Research*. Cambridge, MA:Harvard University Press, 1984.
5. Longnecker MP, Berlin JA, Orza MJ, Chalmers TC. A meta-analysis of alcohol consumption in relation to risk of breast cancer. *JAMA* 260:652-656 (1988).
6. Sugita M, Kanamori M, Izuno T, and Miyakawa M. Estimating a summarized odds ratio whilst eliminating publication bias in meta-analysis. *Jpn J Clin Oncol* 22:354-358 (1992).
7. Dersimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clin Trials* 7:177-188 (1986).
8. Greenland S, Longnecker MP. Methods for trend estimation from summarized dose response data with applications to meta-analysis. *Am J Epidemiol* 135:1301-1309 (1992).
9. Morris RD, Audet AM, Angelillo IF, Chalmers TC, Mosteller F. Chlorination, chlorination by-products, and cancer: a meta-analysis. *Am J Pub Health* 82:955-963 (1992).
10. Romieu I, Berlin JA, Colditz G. Oral contraceptives and breast cancer: review and meta-analysis. *Cancer* 6:2253-2263 (1990).
11. Greenland S. Quantitative methods in the review of epidemiologic literature. *Epidemiol Rev* 9:1-30 (1987).
12. Licciardone JC, Brownson RC, Chang JC, Wilkins JR. Uterine cervical cancer risk in cigarette smokers: a meta-analytic study. *Am J Prev Med* 6:274-281 (1990).
13. Spitzer WO, Lawrence V, Dales R, Hill G, Archer MC, Clark P, Abenhaim L, Hardy J, Sampalis J, Pinfold SP, Morgan PP. Links between passive smoking and disease: a best evidence synthesis. *Clin Invest Med* 13:17-42 (1990).
14. Berlin JA, Colditz GA. A meta-analysis of physical activity in the prevention of coronary heart disease. *Am J Epidemiol* 132:612-628; 1990.