 International Journal of Epidemiology 1999;**28**:1–9

LEADING ARTICLE

Traditional reviews, meta-analyses and pooled analyses in epidemiology

Maria Blettner,[a,b] Willi Sauerbrei,[c] Brigitte Schlehofer,[a] Thomas Scheuchenpflug[a] and Christine Friedenreich[d]

Background	The use of review articles and meta-analysis has become an important part of epidemiological research, mainly for reconciling previously conducted studies that have inconsistent results. Numerous methodologic issues particularly with respect to biases and the use of meta-analysis are still controversial.
Methods	Four methods summarizing data from epidemiological studies are described. The rationale for meta-analysis and the statistical methods used are outlined. The strengths and limitations of these methods are compared particularly with respect to their ability to investigate heterogeneity between studies and to provide quantitative risk estimation.
Results	Meta-analyses from published data are in general insufficient to calculate a pooled estimate since published estimates are based on heterogeneous populations, different study designs and mainly different statistical models. More reliable results can be expected if individual data are available for a pooled analysis, although some heterogeneity still remains. Large prospective planned meta-analysis of multicentre studies would be preferable to investigate small risk factors, however this type of meta-analysis is expensive and time-consuming.
Conclusion	For a full assessment of risk factors with a high prevalence in the general population, pooling of data will become increasingly important. Future research needs to focus on the deficiencies of review methods, in particular, the errors and biases that can be produced when studies are combined that have used different designs, methods and analytic models.

Keywords	Epidemiological methods, meta-analysis, pooled analysis, reviews, risk factors
Accepted	9 July 1998

Much attention has been given in recent years to meta-analysis in medical research, however, numerous methodologic issues particularly with respect to biases and the use of meta-analysis are still controversial.[1–4] One of the issues is whether meta-analysis from published data is sufficient for such an exercise or whether individual patient data are necessary. The latter type of meta-analyses has been performed for several clinical and epidemiological data sets and added considerably to our knowledge of the use of some therapies[5,6] as well as to the role of some aetiologic factors, as for example in the recently published analysis on breast cancer and oral contraceptive use.[7] The Cochrane Collaboration is an important outcome of this discussion.[8] The use of meta-analysis of observational epidemiological studies has been increasing recently, however, its use is less accepted than in the area of clinical trials. Numerous discussions and publications have occurred regarding the merits of meta-analysis of epidemiological studies.[9–16] A series of papers has been published in the *British Medical Journal* mainly concerning meta-analysis of clinical data, but many of the aspects are also relevant for a meta-analysis using observational data.[17–20] Some authors consider a meta-analysis the best possible use of all available data for assessing risk factors and their associations with disease outcomes while others regard the results of meta-analyses sceptically and question whether they add to scientific knowledge. The majority of the controversy centres on

[a] German Cancer Research Center, Division of Epidemiology, Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany.

[b] International Agency for Research on Cancer, 150, cours Albert-Thomas, F-69372 Lyon Cedex 08, France.

[c] Institute for Medical Biometry and Medical Informatics, University Freiburg, Stefan-Meier-Str. 26, D-79104 Freiburg, Germany.

[d] Alberta Cancer Board, Division of Epidemiology, Prevention and Screening, 1331–29 St NW, Calgary, Alberta T2N 4N2, Canada.

meta-analyses of published data, however, some of the critical arguments also apply to meta-analyses of individual patient data.

The use of review articles and meta-analysis has become an important part of epidemiological research. One of the main reasons for their increased use in summarizing the evidence in a particular area is the need to assess risks that are small but that may have large public interest or have important implications for public health, such as passive smoking and lung cancer, the use of oral contraceptives and breast cancer or low-dose radiation exposure and leukaemia.

In this paper we will discuss four different methods for summarizing the evidence. These are:

Type I Qualitative summary, the narrative review article.
Type II Quantitative summary of published data (usually called meta-analysis).
Type III Re-analysis of individual data based on primary studies. This is sometimes called meta-analysis, but in epidemiology the term 'pooled analysis' is more often used.
Type IV Prospectively planned, pooled analysis of several studies, where pooling is already a part of the protocol. Data collection procedures, definition of variables, questions and hypotheses are standardized for the individual studies.

Type IV is different in several respects to a classical multicentre therapy study since the single studies are analysed and published separately. In many situations, the design of the studies is slightly different because of local or regional circumstances.

In this paper we summarize and compare the different types of reviews in epidemiological research and evaluate their usefulness for assessing risk factors. First, we give the major reasons and some general rules for conducting a meta-analysis. Then we describe the similarities and differences between the various types. We compare the advantages and limitations of the four types and illustrate our arguments with an example. Finally, we discuss the need for further research to establish standards for performing meta-analysis in epidemiology.

Rationale for Meta-analyses

The main reason for conducting a review, a meta-analysis or a pooled analysis is to reconcile previously conducted studies that have inconsistent results. Such a situation may arise when the sample sizes of individual studies are too small to find stable results or if the results from single studies vary considerably. Meta-analyses and reviews are mainly used to assess weak risk factors that have a large public health impact (such as passive smoking, the exposure to low-level ionizing radiation, use of contraceptives, exposure to electromagnetic fields or to indoor radon). In contrast to qualitative reviews, meta-analyses are mainly performed to obtain a combined estimator of the quantitative effect of the risk factor such as standardized mortality ratio (SMR) or the relative risk (RR). Some meta-analyses are also used to investigate more complex dose-response functions. The majority of meta-analyses or pooled analyses conducted thus far examined dichotomous (or categorical) risk factors. Briefly, the main reasons for conducting a meta-analysis or a review are:

1) To assess qualitatively whether a factor has to be considered as a risk factor.
2) To provide more precise effect estimates and increased statistical power and to analyse dose-response relations.
3) To investigate the heterogeneity between different studies.
4) To generalize results of single studies.
5) To investigate rare exposures and interactions.
6) To investigate risks associated with rare diseases.

Characterization and Limitations of the Four Types

Review (Type I)

Traditional narrative reviews provide a qualitative but not a quantitative assessment of published results. They are influenced by publication bias[21,22] and file-drawer problem.[23] If there is not an *a priori* strict protocol for the review, narrative reviews are only a subjective judgement of the included studies. However, if they are carefully done, they can give quite an extensive overview of the current state of the research within a short time frame and at low cost. Methodologic guidelines for reviewing papers have recently been published and may add to the quality of reviews.[16]

Meta-analysis from published data (Type II)

An important goal of these studies is to calculate a quantitative pooled estimate of the effect of interest. This type of review can be performed from published data without the cooperation and even without the agreement of other study groups or study coordinators. However, attempts should be made to obtain additional information from study coordinators when necessary. Meta-analysis of published papers has several severe limitations.[13] One limitation is that publication bias is particularly important in epidemiological research since some analyses may be done in a very explorative way and may be only published selectively. As mainly unexpected significant results may be selected for publication, an overestimation of the risk estimate is likely. An additional problem is that studies may differ considerably in their designs, data collection methods and the definition of the exposure and confounder variables. A special dilemma arises when separate studies adjust for different confounding factors. No systematic investigation has been performed to determine whether the simple (crude) estimates or 'best estimates' should be used for combining results of individual studies. We believe that a pooled estimate should not be published if the heterogeneity between studies is high. In many publications of meta-analysis the problem of heterogeneity is not sufficiently assessed. A pooled estimate is often published even when strong—statistically significant—heterogeneity between study results was found (and reported).

Meta-analyses with individual data (Type III)

Some of the problems that arise with Type II meta-analysis are avoidable if individual data from all studies are available. Several recent publications have used this approach.[7,24] Publication bias may be reduced in comparison to Type I and II meta-analyses as even unpublished data can be included. The cooperation between different study centres and research coordinators may also help to identify further studies that are only known to some investigators. With individual data, statistical re-analysis can be performed. This includes some inclusion criteria for all studies, a unified definition of the variables, and new statistical

modelling. With a large sample size the effect of rare exposures can also be examined, which is often not possible in single studies. New hypotheses may be formulated with these types of pooled analyses and specific subgroups, such as age groups, may also be analysed.

A major obstacle for this kind of meta-analysis, is the fact that it is more expensive, time-consuming, requires close cooperation between study coordinators and several years to complete.[4,7,25] In addition, improvement of data quality is not possible, however, some errors in the data or analysis can be partially eliminated. Furthermore, adjustment for confounding variables that have been delineated since the original studies were performed can now be done if those covariables were originally collected. Differences between the study results can be actively discussed between study coordinators and reasons for these differences can be elucidated. In general, it is possible to estimate risk coefficients from pooled data and to calculate a variance and confidence interval.

Prospectively planned pooled analysis (Type IV)

This type of analysis has not been called a meta-analysis, despite the fact that it has several aspects in common with Type III meta-analyses. Several examples of large international case-control studies[26] and occupational cohort studies have used this approach.[27] The main difference from Type III meta-analysis is a joint planning of the data collection and analysis to avoid large differences between the studies. The experience of many researchers is used in the planning to ensure comparability in designs, data collection, data analysis and reporting across all centres. In contrast to multicentre clinical trials, more heterogeneity in the individual study centres will still exist. This heterogeneity arises from differences in populations (e.g. race is not a confounder factor in Germany but is in the USA) or differences in design (e.g. no methods of random population sampling exist in the USA, no overall cancer registration in Germany). An assessment of any remaining sources of heterogeneity is possible. The costs for Type IV meta-analyses are very high. The planning can be substantial, even difficult and the time incurred can be long. Another disadvantage of this type of meta-analysis is that errors in the design of single studies are multiplied. An alternative approach is to have individual studies but a joint core protocol for questions of common interest. This design allows individual coordinators to investigate specific hypotheses and permits some variation across individual studies. Moreover, once such a meta-analysis has been performed it will be more difficult to justify a new individual study with the same topic.

Methods for an Overview

All types of overview—whether quantitative or qualitative—have some steps in common that should be followed in their planning and conducting. Each individual type has some aspects of the conduct that are different and these are described later.

Steps in performing a meta-analysis

Each type of overview needs a clear study protocol that describes the research question and the design, including how studies are identified and selected, the statistical methods to use and how the results will be reported. This protocol should also include the exact definition of the disease of interest, the risk factors and the potential confounding variables that have to be considered. A main component of the protocol is the exact definition of the inclusion criteria for single studies. As described by Friedenreich,[28] the following steps are needed for a meta-analysis/pooled analysis:

1) Define a clear and focused topic for the review.
2) Locate all studies (published and unpublished) that are relevant to the topic.
3) Select all studies that are relevant according to the explicit inclusion criteria.
4) Abstract necessary information from the published papers or obtain the primary data from the original investigators. Meta-analysis of published data may also include contacting the original project leaders to obtain data or information that have not been published in sufficient detail. For a pooled analysis, agreement to use the original data is needed.
5) Tabulation of relevant elements of each study, including sample size, assessment procedures, available variables, study design, publication year, performing year, geographical setting, etc.
6) Define protocol for the analysis of all studies and estimate the study-specific effects (relative risks adjusted for relevant confounder variables).
7) Investigate the homogeneity of study-specific effects and determine whether these effects can be combined to perform a pooled analysis.
8) Presentation of published results, e.g. graphically.
9) Investigate and reduce (if possible) the heterogeneity between studies.
10) Decide about remaining heterogeneity components: Coping with different designs, study types, confounders, etc.
11) Estimate a pooled effect with adequate statistical methods if the studies are efficiently homogeneous.
12) Conduct a sensitivity analysis.

For traditional reviews, only the first two steps with the qualitative assessment are done. For a meta-analysis from published data, data abstraction will be done from publications, however, if data are needed that are not given in the publications, contact with the project manager of the studies should be made. For Type III and IV meta-analysis, all steps are necessary. For Type IV meta-analysis, the two first steps involve writing a study protocol for the individual and the pooled studies. This protocol should include a joint questionnaire (with questions used for the pooled analysis and additional questions added by single investigators), a common study design and permitted deviations from the study design for local reasons. As the data collection is conducted in each centre, an additional step required for Type III and IV meta-analysis is the pooling of all data with subsequent data processing and management.

Statistical analysis

The statistical analysis of aggregated data from published studies was first developed in the fields of psychology and education.[29,30] These methods have been adopted since the mid-1980s in medicine primarily for randomized clinical trials but also used for observational epidemiological studies. We will give a brief outline of some issues of the analysis; further details will be the subject of a subsequent publication.

Single study results

The first step of the analysis is the description of the characteristics and the results of each study. Tabulations and simple graphical methods should be employed to visualize the results of the single studies. Plotting the odds ratios and their confidence intervals (so-called forest plot) is a simple way to spot obvious difference between the study results. The radial plot[31] is a more sophisticated means of investigating the heterogeneity and the contribution of each study to the overall estimate.

Heterogeneity

The investigation of heterogeneity between the different studies is a main task in each review or meta-analysis.[32] For the quantitative assessment of heterogeneity, several statistical tests are available.[33,34] A major limitation of formal heterogeneity tests is, however, their low statistical power to detect any heterogeneity present. Therefore, heterogeneity should be investigated informally, e.g. by comparing results from studies with different designs, within different geographical regions. In addition, graphical methods should be used to visualize heterogeneity, such as plots with single studies grouped or ordered according to special covariables (e.g. type of study, publication time), funnel plots to indicate publication bias, and radial plots to indicate studies which are very influential on this heterogeneity (see example).

In meta-analysis of published data (Type II) only sensitivity analysis can be performed to investigate the degree of heterogeneity. However, if individual data are available (Type III, IV), the sources of heterogeneity can be investigated in detail and some degree of heterogeneity can be reduced by using the same statistical model for all single studies. For Type IV meta-analysis, the analytic strategy and definitions can be determined *a priori* for all of the individual studies. Hence, an identical multiple regression analysis can be used in each centre. This approach avoids heterogeneity that could be introduced by inconsistent modelling strategies (e.g. different variable selection and definition, confounder adjustment).

Pooling

Whether the data can be pooled should be decided after investigating the homogeneity of the study results. If the results vary substantially, no pooled estimate should be presented or only estimates for selected subgroups should be calculated (e.g. combining results from case-control studies only). Methods for pooling depend on the data available. In general, a two-step procedure has to be applied. First, the risk estimates and variances from each study have to be abstracted (Type II) or calculated (Type III, IV). Then, a combined estimate is obtained as a (variance based) weighted average of the individual estimates. The methods for pooling based on the 2×2 table include the approaches by Mantel-Haenszel and Peto.[34,35] If data are not available in a 2×2 table, but as estimates from a more complex model (such as an adjusted relative risk estimate), the Woolf and DerSimonian-Laird approach can be adopted using the estimates and their (published or calculated) variances resulting from the regression model.[36] For these methods, variance estimates of the pooled estimator are available and allow the calculation of confidence intervals.

Two different statistical models can be used to estimate combined risks. In the *fixed effects model*, it is assumed that the underlying true exposure effect in each study is the same. The overall variation and, therefore, the confidence intervals reflect the random variation within each study but not any potential heterogeneity between the studies. If individual data are available, the pooled estimator and its variance can be obtained using regression models by incorporating an additional dummy variable for each centre. This analysis can be done with either Poisson regression or a (conditional) logistic regression. *Random effects models* incorporate variation between the studies. It is assumed that each study has its own (true) exposure effect and that there is a random distribution of these true exposure effects around a *central* effect. The observed effects from the different studies are used to estimate this distribution. In other words, the random effects model allows non-homogeneity between the effects of different studies.

The most common random effects model to combine the single estimates is the method of moments approach given by DerSimonian and Laird.[36] The important difference is that for this model study-specific weights are calculated as a sum of the variance within the studies and a term for the variance between the studies, τ^2. The between-study variance τ^2 can also be interpreted as a measure for the heterogeneity between studies.

Comparison between fixed effects and random effects models

- Random effects methods yield (in general) larger variance and confidence intervals than fixed effects models because a between-study component τ^2 is added to the variance.
- If the heterogeneity between the studies is large, τ^2 will dominate the weights and all studies will be weighted more equally (in random effects model weight decreases for larger studies compared to the fixed effects model)
- A major critique of the random effects model is that it is not sufficient to 'explain' the heterogeneity between studies, since the random effect merely quantifies unexplained variation by estimating it.[37] Heterogeneity between studies should yield careful investigation of the sources of the differences. If a sufficient number of different studies are available, further analyses, such as 'meta-regression', may be used to examine the sources of heterogeneity.[12,38]

If individual data are available, the fixed effects estimate can be calculated from a regression model with dummy variables. So far, there is no comparable approach available for the random effects model. Here, the two-step procedure is used even with individual data available.[24]

Several other methods have been proposed to estimate the overall effect based on maximum-likelihood models or on Bayesian methods.[39,40] Recent investigations have demonstrated that, for practical purposes, the differences between these methods are not very large. So far, only rather sophisticated software is available for these approaches.[41]

Sensitivity analysis

An important method for investigating heterogeneity is sensitivity analysis, e.g. to calculate a pooled estimate only for subgroups of studies (according to study type, quality of the study, period of publication, etc.) to investigate variations of the odds ratio. An extension of this method is meta-regression as proposed by Greenland,[38] however, meta-regression has limitations because of the small number of studies included in most meta-analyses.

Table 1 Comparison of methods for different literature review methods

Requirement for the review method	Type I Review	Type II Meta-analysis from published data	Type III Meta-analysis with individual data	Type IV Prospective planned pooled analysis
Planning and literature search				
Protocol	+?	+	++	++
Inclusion/Exclusion criteria	+	+	++	++
Systematic literature search (incl. Abstracts, Proceedings)	+?	+	++	*
Obtaining additional information from single studies not published	–	+?	+	*
Evaluation of sources of errors and bias				
Investigation of sources of bias	+?	+?	++	++
Evaluation of validity of individual studies	–	+?	++	*
Control of data collection	–	–	+?	++
Adjustment of inclusion criteria for individuals	–	–	+	++
Assessment and control of statistical analysis	–	–	++	++
Estimation of publication bias	–	?	+	*
Comparability of single studies				
Standardized study design	+?	+	+	++
Standardized assessment of risk factors	–	–	–	+
Standardized definition of exposure and confounder variables (categories)	–	–	+?	++
Standardized adjustment for confounder variables	–	–	+?	++
Statistical analysis				
Quantitative estimate for the effect	–	+?	++	++
Improvement of the precision of effects measured	–	?	+	++
Estimator for dose-response relationship	–	–	+?	++
Estimator for risk in subgroups	–	?	+	+
Increase of statistical power	–	+?	++	++
Evaluation of interactions and confounder effects	–	–	+	++
Evaluation of sources of heterogeneity	?	+?	++	++
Sensitivity analyses	–	+	++	++
Reproducibility of methods	–	–	+?	+?
General aspects				
Description of state of research	+	+	+	*
New research questions	++	++	+	+
Improvement of the quality of further studies	+	+	+	+
Time and costs for the study	very low	low	high	very high

++ Possible in principle and (almost) always done.

\+ Possible in principle and often done.

+? Often not possible or not done.

? Only possible in exceptional cases or useful.

– Never possible.

* Less relevant.

Comparison and Assessment of the Four Types of Meta-analysis

A comparison of different literature review methods is outlined in Table 1.

Design, conduct and literature search

For each type of review, the hypothesis, question and conduct should be summarized and defined in a strict protocol in which clear inclusion and exclusion criteria for the studies and the details of the literature search are described. This component of the review process is important for each study type, but is especially needed if quantitative results are required. It should also be decided whether and which data will be required from the investigators of the individual studies.

An important problem of meta-analysis is publication bias. This bias has received a lot of attention particularly in the area of clinical trials.[17–19] Publication bias occurs when studies that have non-significant or negative results are published less frequently than positive studies. For randomized clinical trials, it has been shown that even with a computer-aided literature search only some of the relevant studies will be identified.[42] For

epidemiological observational studies additional problems exist. Very often a large number of variables will be collected in questionnaires as potential confounders. If one or several of these potential confounders yield significant or important results, they may be published in additional papers, which have often not been planned in advance. If these confounders yield expected or negative results they are usually not published. Some regional studies may not be published in international journals and are seldom found by a literature search for a meta-analysis. This situation is mainly a problem for 'replication studies' being conducted more frequently in epidemiology. These studies are usually not published in international journals as they do not add anything new to existing knowledge.

Inclusion criteria, data collection methods and statistical analyses cannot be changed if published data are used for a meta-analysis. In many situations, it is even difficult to determine exactly what has been done in the study based on the published information. The methods section in many papers is often short and critical evaluation is not always possible. Errors in the original work cannot be corrected or checked and may yield to bias in the results of the meta-analysis. For Type III meta-analysis, the inclusion criteria for the single studies can be modified (for different age groups, tumour sites, latency times, etc.). They can also be redefined and checked. It is also possible to evaluate or adopt the statistical analysis for Type III and Type IV meta-analysis. Possible sources of a systematic bias can be eliminated if a detailed statistical analysis of the single studies can be done. The evaluation of possible bias attributable to lack of control for confounding is only possible in Type III and Type IV meta-analysis.

Validation of comparability of the single studies

Since many study designs are possible in epidemiological research it is necessary to evaluate the comparability of the single studies before conducting a review. This evaluation can be made partly from published data if enough detailed information is available in the papers. If individual data are available, an analysis of the single studies in one common model is possible. A major reason for different results across studies is that different statistical methods/models have been used. Hence, heterogeneity can be significantly reduced in a pooled analysis by using the same model for all studies. A pooled analysis is only possible if similar data are available from all studies and are provided to the investigator of the pooled analysis. The investigation of study-specific heterogeneity can be done only to some extent in Type II meta-analysis, mainly by a sensitivity analysis.

Quantitative risk estimation

Narrative reviews (Type I) are not designed to give a quantitative estimate of the effect of risk factors or to describe a dose-response relation. They only allow a descriptive comparison of the results of available research. All other types of meta-analysis allow the calculation of a pooled risk estimate, provided that the data are sufficiently homogeneous. However, calculating an overall risk estimate is not always possible if different statistical models were used in the original studies. An improvement in the precision of the risk estimate is not always achieved. Pooling decreases the variation caused by random error (increasing the sample size) but does not eliminate any bias (systematic error). Indeed, in some situations the pooled estimate is less precise than estimates of the included single studies as was shown by Gilbert[43] in the radiation leukaemia studies.

A less precise pooled estimate is likely if a crude categorization (e.g. 2×2 tables) is needed and no other data have been published. Bias may increase if different methods have been used in the individual studies to assess and control confounding but cannot be forwarded to the pooled estimator. A better estimate of the effect is only possible when individual exposure data are available (Type III and IV). Furthermore, to estimate a dose-response relation, individual subject data are needed, particularly if individual studies have used different exposure categories. The dose-response relation cannot be estimated with sufficient precision with published data only. Likewise an investigation of interaction and confounding requires individual subject data. With published data, the measurement methods and exposure categorization are usually different across studies. Hence, it is misleading to use 2×2 tables for the pooled estimate, if confounders play an important role. Prospective multicentre meta-analyses (Type IV) have the advantage that design, measurement methods, exposure assessment and the definition of all variables can be agreed upon before data collection. Consequently, the data can be more easily combined at a later stage. It should also be noted that subgroup analysis, which is often a goal of a planned meta-analysis, can only be performed if the data are published with sufficient detail.

Investigation of consistency

To investigate whether the results are consistent across studies, published data can be used for a narrative review (Type I) as well as for Type II meta-analysis. The method of choice is a sensitivity analysis in which the results are compared for different strata, e.g. same study design, same risk assessment method, same statistical analysis. Possible means for a sensitivity analysis include: exclusion of studies with particular heterogeneous results (outliers), conducting a separate analysis for case-control and for cohort studies or combining studies from similar geographical areas or for similar time periods. However, it is only possible in Type III and Type IV meta-analyses to change exposure and confounder variable categories or use the same or similar confounder and examine interaction in a multivariate statistical model and perform detailed analysis of important subgroups. A valid judgement of the consistency of results in complex questions requires such a detailed statistical analysis.

Many authors have pointed out that investigating heterogeneity is the most important aspect of meta-analysis (e.g. Thompson[32]). In our opinion such an investigation is particularly important for meta-analyses of epidemiological data. Statistical methods to investigate heterogeneity can be based on aggregated data. However, statistical tests have low power and may not be able to detect heterogeneity between studies. Type III meta-analysis allows different strategies to be used to eliminate differences and at least to give results in a unified way. It should be noted that it is sometimes difficult to compare results from different epidemiological studies since different data presentation methods are used across publication. Even in a single study different strategies for modelling can yield rather different results.[44] Therefore, for a meaningful meta-analysis it is necessary to eliminate this source of heterogeneity. Such a comparison is only possible if the same risk factors and confounding factors are available from all single studies.

Example: Oral Contraceptive Use and Breast Cancer

The role of oral contraceptive (OC) use as a risk factor for breast cancer remains a controversial issue despite many years of research. Several epidemiological studies of this association have been conducted. These studies have yielded inconsistent results; in most studies, odds ratios varied around one with large confidence intervals. Previously conducted studies have also varied in the type of exposure information obtained. Information on OC use ranged from 'ever-never' use to more detailed information on duration and frequency of use, dosage, age at first use, type of OC, etc. These differences in exposure data make it difficult to compare results across studies. In addition, different analytic models were used since different confounders were available for the analysis. Some, but not all, studies adjusted for variables such as age, age at first birth, parity, breastfeeding, age at menopause or other confounders. As the prevalence of OC use is high in many societies, a quantitative risk estimator is needed and important for personal and health policy decisions. Since large methodologic differences exist across studies, summarizing the results in an overall quantitative estimate is difficult.

Several reviews and meta-analyses have been published.[45,46] We will here compare one Type II meta-analysis[47] and a pooled analysis including individual data (Type III) from over 50 epidemiological studies[7] to illustrate some of the issues discussed above. Rushton and Jones[47] included 27 case-control studies that have been published between 1980 and 1989, however, they included only subsets of these studies for specific research questions. For example, only 20 of the studies included age as a confounder, and only 24 studies had at least some information on the duration of OC use. Only a few of the 27 studies measured all the important confounder variables, such as parity. In this meta-analysis, the authors dichotomized all variables and estimated an overall pooled odds ratio (RR 1.10; 95% CI : 1.04–1.17). They also estimated odds ratios for different subgroups of the data, including by age (RR 1.16; 95% CI : 1.07–1.25 for women <45 years), by parity (RR 1.21; 95% CI : 0.99–1.47 for women without children), and by length of OC use (RR 1.27; 95% CI : 1.12–1.44 for duration >8 years). Rushton and Jones found significant heterogeneity across individual studies that could not be explained by the design factors that they investigated.

The pooled analysis of the Collaborative Group on Hormonal Factors in Breast Cancer[7] used individual data from 54 studies (case-control and cohort studies), including data from more than 50 000 breast cancer cases and 100 000 controls. The authors estimate that these comprise about 90% of all studies that have been performed on this topic by the time of their analysis. The pooled analysis used the same statistical models, identical definition of risk factors (OC use) and inclusion criteria. Also, the same confounder variables were considered (study centre, age at first diagnosis, parity, age at first birth, age at menopause, family history of breast cancer, duration of pill use, time between last use and breast cancer diagnosis, age at first and age at last use). For all data combined, a small increase in risk of breast cancer was observed for women during OC use (RR 1.24; 95% CI : 1.15–1.33). They also found that 10 years after the exposure to OC use stopped, the risk was no longer increased (RR 1.01; 95% CI : 0.96–1.05) compared to women who never used OC. The authors also investigated several types of interaction between OC use, family history, parity, etc. These analyses were possible since individual data were available and could be re-analysed.

The results from the studies are rather homogeneous (after using the same model for each analysis), but there is some remaining heterogeneity. This is shown in the forest plot (Figure 1) where the study-specific risks and their confidence intervals are plotted,

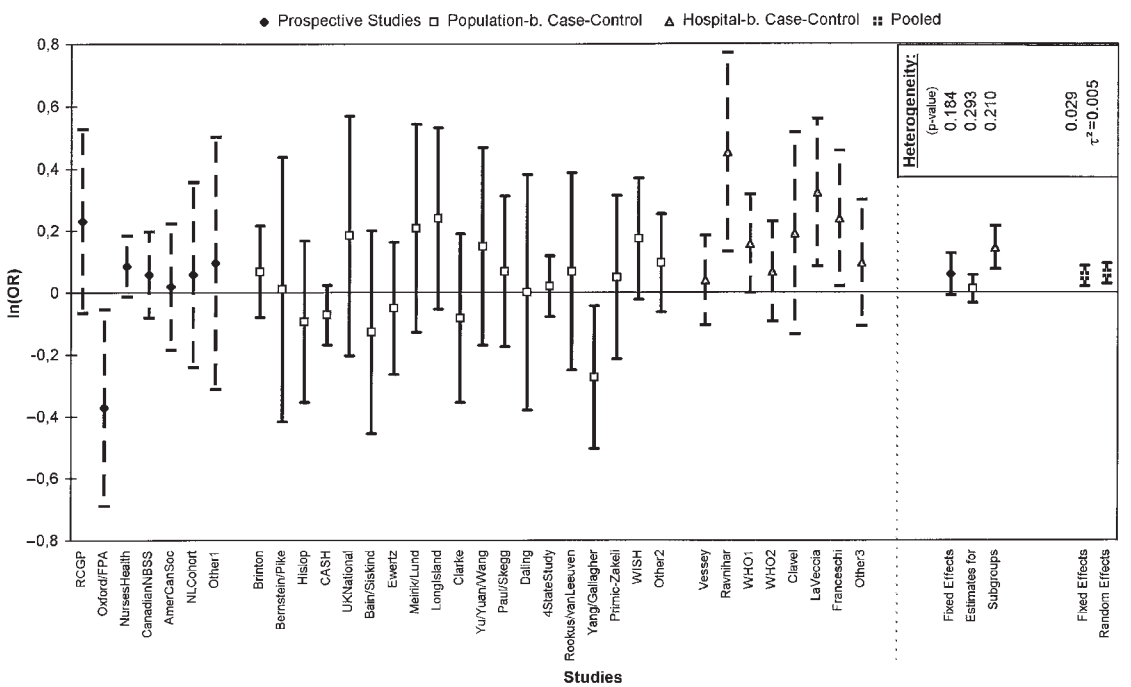

Figure 1 Relative risk of breast cancer in ever-users compared with never-users of combined oral contraceptives—forest plot

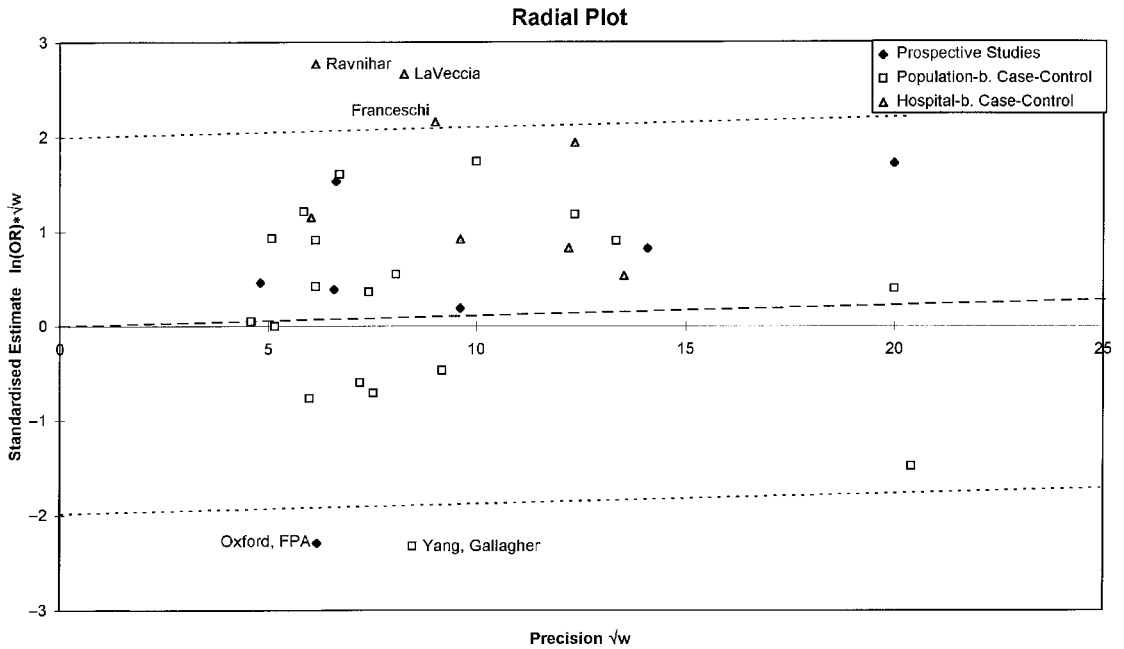

Figure 2 Relative risk of breast cancer in ever-users compared with never-users of combined oral contraceptives—radial plot

grouped by different study types. Results differ slightly between case-control studies with different controls (population versus hospital controls) and between case-control studies and cohort studies. The radial plot (Figure 2, see Galbraith[31] for details) indicates some 'outlying' studies, especially hospital-based case-control studies. Heterogeneity cannot be explained by differences in study design since the analysis was performed similarly across the different studies. However, the pooled analysis was restricted to variables that have been collected in the individual studies.

From this it is clear that pooled analysis of individual data is preferable to meta-analysis from published data. The latter may in general be insufficient for calculating pooled estimates but can give a structured overview of available published literature.

Recommendations for Future Research and Conclusion

This paper has described different review methods for observational epidemiological studies. It is clear from this discussion that all available data and information will be needed for full assessment of weak risk factors with a high prevalence and that pooled analysis will become increasingly important. There is an immense need to evaluate the risk associated with many exposures, particularly where the previously conducted epidemiological studies have provided inconsistent results. To assess risks, summaries of available and published data need to be conducted. As discussed, a major impediment for meta-analysis of epidemiological data is the heterogeneity across studies in their design, data collection methods and analyses performed. Meta-analyses using published data are, therefore, restricted and seldom useful to produce a valid quantitative estimate or to investigate exposure relations such as dose-response. Nonetheless, as we have discussed, a Type II meta-analysis of published data may be more reproducible than a Type I qualitative review. Type II meta-analysis also has the distinct advantage of being less expensive and time-consuming to perform than meta-analysis with individual data (Type III, IV). Consequently, many researchers will continue to perform meta-analyses and, on the other hand, more and more public health regulators and decision-makers may rely on or require their results for decision-making.

Future research in this field needs to focus on the deficiencies of each type of review method in order to improve this methodology. In particular, the errors and biases that can be produced when studies are combined that have used different design, methods and analytic models need to be addressed. Other areas for consideration in future studies include the influence of different baseline risks, the different quality and type of the exposure measurements made, the methods for pooling studies that have measured different confounding variables. Simulation studies or more theoretical approaches may be required to assess all of those factors and their influence on the results obtained from different types of meta-analysis. Statistical methods for pooling data from different sources have to be refined and new approaches, such as Bayesian methods may need to be implemented. More practical experience is also needed with these different approaches. Despite advances in computer technology, pooled analysis with individual data remain difficult when large numbers of variables and complex statistical models are run. The methods to use for the proper conduct and reporting of meta-analyses of published or individual data from epidemiological studies still require more refining, consensus and widespread use by the scientific community. Significant progress has been made in the systematic approaches for meta-analyses of clinical studies. It is now increasingly important that these methods for epidemiological studies have equally rigorous standards since more public health decisions will be relying,

to a greater extent, on the results of meta-analyses. Hence, the epidemiological community must ensure that the validity, reliability and overall quality of these methods is improved and implemented by all who are using these methods.

References

[1] Chalmers TC. Problems induced by meta-analyses. *Stat Med* 1991;**10:** 971–80.

[2] Chalmers TC, Lau J. Meta-analytic stimulus for changes in clinical trials. *Stat Methods Med Res* 1993;**2:**161–72.

[3] Thompson SG, Pocock SJ. Can meta-analyses be trusted? *Lancet* 1991;**338:**1127–30.

[4] Stewart LA, Parmar MKB. Meta-analysis of the literature or of individual patient data: is there a difference? *Lancet* 1993;**341:**418–22.

[5] EBCTCG—Early Breast Cancer Trialists' Collaborative Group. Systemic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy. 133 randomised trials involving 31,000 recurrences and 24,000 deaths among 75,000 women. *Lancet* 1992;**339:**1–15.

[6] AOCTG—Advanced Ovarian Cancer Trialists' Group. Chemotherapy in advanced ovarian cancer: an overview of randomised clinical trials. *Br Med J* 1991;**303:**884–93.

[7] CGHFBC—Collaborative Group On Hormonal Factors In Breast Cancer. Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53 297 women with breast cancer and 100 239 women without breast cancer from 54 epidemiological studies. *Lancet* 1996;**347:**1713–27.

[8] Chalmers I. The Cochrane Collaboration: preparing, maintaining, and disseminating systematic reviews of the effects of health care. *Ann NY Acad Sci* 1993;**703:**156–63.

[9] Olkin I. Invited Commentary: Re: "A critical look at some popular meta-analytic methods". *Am J Epidemiol* 1994;**140:**297–99.

[10] Beral V. "The practice of meta-analysis": discussion. Meta-analysis of observational studies: a case study of work in progress. *J Clin Epidemiol* 1995;**48:**165–66.

[11] Berlin JA. Invited Commentary: benefits of heterogeneity in meta-analysis of data from epidemiologic studies. *Am J Epidemiol* 1995; **142:**383–87.

[12] Greenland S. Invited Commentary: a critical look at some popular meta-analytic methods. *Am J Epidemiol* 1994;**140:**290–96.

[13] Shapiro S. Meta-analysis/Shmeta-analysis. *Am J Epidemiol* 1994a;**140:** 771–78.

[14] Shapiro S. Is there is or is there ain't no baby?: Dr. Shapiro replies to Drs. Petitti and Greenland. *Am J Epidemiol* 1994b;**140:**788–91.

[15] Feinstein AR. Meta-analysis: statistical alchemy for the 21st century. *J Clin Epidemiol* 1995;**48:**71–79.

[16] Weed DL. Methodologic guidelines for review papers. *J Natl Cancer Inst* 1997;**89:**6–7.

[17] Egger M, Davey Smith G. Misleading meta-analysis. *Br Med J* 1995; **310:**752–54.

[18] Egger M, Davey Smith G. Meta-Analysis. Potentials and promise. *Br Med J* 1997;**315:**1371–74.

[19] Egger M, Davey Smith G. Bias in location and selection of studies. *Br Med J* 1998a;**316:**61–66.

[20] Egger M, Schneider M, Davey Smith G. Spurious precision? Meta-analysis of observational studies. *Br Med J* 1998b;**316:**140–44.

[21] Dickersin K. The existence of publication bias and risk factors for its occurrence. *JAMA* 1990;**263:**1385–89.

[22] Dickersin K. How important is publication bias? A synthesis of available data. *AIDS Educ Prev* 1997;**9(Suppl A):**15–21.

[23] Rosenthal R. The "file-drawer problem" and tolerance for null results. *Psychol Bull* 1979;**86:**638–41.

[24] Lubin JH, Boice, JD Jr, Edling C *et al.* Radon-exposed underground miners and inverse dose-rate (protraction enhancement) effects. *Health Phys* 1995;**69:**494–500.

[25] Stewart LA, Clarke MJ. on behalf of the Cochrane Working Group on Meta-Analysis Using Individual Patient Data. Practical methodology of meta-analyses (overviews) using updated individual patient data. *Stat Med* 1995;**14:**2057–79.

[26] Preston-Martin S, Pogoda JM, Schlehofer B *et al.* An international case-control study of adult glioma and meningioma: the role of head trauma. *Int J Epidemiol* 1998;**27:**579–86.

[27] Boffetta P, Saracci R, Andersen A *et al.* Cancer mortality among man-made vitreous fiber production workers. *Epidemiology* 1997;**8:**259–68.

[28] Friedenreich CM. Methods for pooled analyses of epidemiologic studies. *Epidemiology* 1993;**4:**295–302.

[29] Glass GV. Integrating findings: the meta-analysis of research. *Rev Res Ed* 1977;**5:**3–8.

[30] Smith ML, Glass GV. Meta-analysis of psychotherapy outcome studies. *Am Psychol* 1977;**32:**752–60.

[31] Galbraith RF. Some applications of radial plots. *JASA* 1994;**89:** 1232–42.

[32] Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. *Br Med J* 1994;**309:**1351–55.

[33] Paul SR, Donner A. A comparison of tests of homogeneity of odds ratios in K 2×2 tables. *Stat Med* 1989;**8:**1455–68.

[34] Petitti DB. *Meta-Analysis, Decision Analysis and Cost-Effectiveness Analysis. Methods for Quantitative Synthesis in Medicine.* New York, Oxford: Oxford University Press, 1994.

[35] Hedges LV, Olkin I. *Statistical Methods for Meta-Analysis.* Orlando: Academic Press, 1985.

[36] DerSimonian R, Laird N. Meta-analysis in clinical trials. *Controlled Clin Trials* 1986;**7:**177–88.

[37] Mengersen KL, Tweedie RL, Biggerstaff BJ. The impact of method choice on meta-analysis. *Austral J Statist* 1995;**37:**19–44.

[38] Greenland, S. Quantitative methods in the review of epidemiologic literature. *Epidem Rev* 1987;**9:**1–30.

[39] DuMouchel W. Bayesian Metaanalysis. In: Berry DA (ed.). *Statistical Methodology in the Pharmaceutical Sciences.* New York: Dekker, 1990, pp.509–29.

[40] Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to random-effects meta-analysis: a comparative study. *Stat Med* 1995; **14:**2685–99.

[41] Spiegelhalter DJ, Thomas A, Best N, Gilks W. BUGS: Bayesian Inference Using Gibbs Sampling (V 0.5, 1996): Cambridge: MRC Biostatistics Unit, Institute of Public Health, (ftp.mrc-bsu.cam.ac.uk).

[42] Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *Br Med J* 1994;**309:**1286–91.

[43] Gilbert ES. Analyses of combined mortality data on workers at the Hanford Site, Oak Ridge National Laboratory and Rocky Flats Nuclear Weapons Plant. *Radiat Res* 1989;**120:**19–35.

[44] Blettner M, Sauerbrei W. Influence of model-building strategies on the results of a case-control study. *Stat Med* 1993;**12:**1325–38.

[45] Prentice RL, Thomas DB. On the epidemiology of oral contraceptives and breast disease. *Adv Cancer Res* 1987;**49:**284–401.

[46] Ollsson H. Oral contraceptives and breast cancer. A review. *Acta Oncol* 1989;**6:**849–63.

[47] Rushton L, Jones DR. Oral contraceptive use and breast cancer risk: a meta-analysis of variations with age at diagnosis, parity and total duration of oral contraceptive use. *Br J Obstet Gynaecol* 1992;**99:**239–46.