Canadian Journal of Nursing Research, 1996, Vol. 28, No. 3, 69-85

The Meta-Analytic Approach to Research Integration

Judee E. Onyskiw

La méthode de recherche quantitative pour rassembler des résultats à partir de la méta-analyse d'une recherche empirique est utile pour intégrer systématiquement les résultats glanés dans des recherches individuelles. La méta-analyse va au-delà de la simple récapitulation et critique des résultats de la recherche; on procède à des analyses statistiques sur les résultats de recherches similaires. Globalement, la méta-analyse représente une immense amélioration par rapport aux méthodes traditionnelles de l'examen de la recherche; elle donne une description plus complète du statut actuel de la recherche dans un domaine et une évaluation plus précise des effets des traitements ou des interventions. Le présent article décrit la méta-analyse appliquée à l'intégration de la recherche. Il présente les avantages que cette méthode offre à l'intégration de la recherche en sciences infirmières et met en lumière quelques questions méthodologiques concernant cette méthode.

A quantitative research method for aggregating findings from empirical research, meta-analysis is useful for systematically integrating findings gleaned from individual studies. Meta-analysis goes beyond the mere summarization and critique of research findings, to conducting statistical analyses on the outcomes of similar studies. Overall, meta-analysis represents a vast improvement over traditional methods of research review, by providing a more thorough description of the current status of research in an area and a more precise estimate of the effects of treatments or interventions. This article describes the meta-analytic approach to research integration, discusses the advantages that it offers for integrating nursing research, and highlights some of the methodological issues surrounding this approach.

Synthesizing and integrating research findings is integral to most research endeavours. Before embarking on a new research project, the investigator must put results of past research into a coherent form, in order to understand the current state of knowledge and to identify areas that require further investigation (Reynolds, Timmerman, Anderson, & Stevenson, 1992). Reviews not only form a critical link between past and future research but are fundamental to the accumulation of knowledge – because knowledge does not accumulate from the results of any single study. It is only when findings from numerous

---

Judee E. Onyskiw, R.N., M.N., is a doctoral candidate in the Faculty of Nursing, University of Alberta, Edmonton. Address correspondence to: Judee E. Onyskiw, Faculty of Nursing, University of Alberta, 3rd Floor, Clinical Sciences Building, Edmonton, Alberta T6G 2G3, Canada. Phone: (403) 492-4635; fax: (403) 492-2551; e-mail: jonyskiw@gpu.srv.ualberta.ca

studies of the same phenomenon are integrated, synthesized, and organized into a rational pattern that knowledge in any discipline develops.

The traditional approach to integrating findings from previous research has been the narrative review. Literature on a topic is collected, reviewed, and summarized based on the criteria deemed important by the reviewer. This approach has been criticized because it lacks formal rules, poses difficulties when the literature base is large, and often leads to conflicting reviews of the same literature by different reviewers (Curlette & Cannella, 1985). Meta-analysis was developed to overcome some of the problems inherent in the narrative review, by increasing the objectivity of the review process. As a quantitative method, it goes beyond the mere summarization and critique of research findings, by providing statistical analyses on the outcomes of similar studies (Pillemer & Light, 1980). This approach can be taken by nurses to meet the challenge of making sense of the growing body of nursing research in order to guide knowledge development and practice.

The purpose of this article is to describe the meta-analytic approach to research integration, discuss the advantages it offers to integrating nursing research, and highlight some of the methodological issues surrounding this approach.

Definition

The term "meta-analysis," which was coined by Gene Glass in 1976, is derived from the Greek prefix "meta," meaning "transcending," and the root, analysis. Glass differentiates meta-analysis from primary analysis, which is the original analysis of data, and from secondary analysis, which is the re-analysis of data either to answer the original research question with more sophisticated statistical techniques or to answer an entirely new research question. Glass defines meta-analysis as "the statistical analysis of a large collection of results from individual studies for the purpose of integrating findings" (1976, p. 3). Individual research studies are the subjects in a meta-analytic study as well as the unit of analysis (Devine, 1990). Since meta-analysis depends on the findings of primary research for its data, it is often referred to as an *analysis of analyses.*

Procedures

A meta-analytic study uses formal procedures to combine the findings of several empirical studies. It incorporates a systematic and objective process that parallels the research process in terms of scientific inquiry,

analogous phrases, and attempts to minimize threats to internal and external validity (McCain, Smith, & Abraham, 1986; Smith & Naftel, 1984). The researcher must carefully document the procedures used and the decisions made in order to allow others to assess the adequacy of the study and to allow for replication. Other researchers employing the same techniques should arrive at the same statistical conclusion (Pillemer & Light, 1980).

A meta-analysis involves the following steps: (1) defining the problem and establishing inclusion criteria for admissible studies, (2) retrieving relevant studies, (3) classifying and coding the study characteristics, (4) converting the outcome measure to a common scale or metric, (5) aggregating the study findings, and (6) interpreting the study findings. While meta-analysis was initially developed as a tool to integrate research from experimental or quasi-experimental designs, mathematical estimations have been developed to allow the technique to be used with descriptive research as well (Hedges & Olkin, 1985; Reynolds et al., 1992).

Defining the Problem and Establishing Inclusion Criteria for Admissible Studies

Clearly defining the research question is the first step in any study. The meta-analysis is driven by a specific research question, clearly formulated in terms of the target population, the treatment or intervention of interest, and the outcome measure. Next, systematic, objective, explicit procedures are developed to guide the retrieval of relevant studies. The systematic and objective nature of meta-analysis minimizes the potential for bias, while making the procedures explicit ensures that the study can be replicated (Petitti, 1994). Criteria are defined to determine which studies will be included in the meta-analysis. These criteria should be specific and the rationale for each criterion should be carefully documented in the study protocol.

Retrieving Relevant Studies

The second step is a thorough search of the literature to locate all studies pertinent to the topic. To avoid publication bias, it is critical that both published and unpublished studies be included (Dickerson, 1990; Smith, 1980). Cooper (1982) describes five approaches to locating studies: (1) computer searches, (2) abstracting services, (3) the descendancy approach, (4) the ancestry approach, and (5) the invisible college approach. In the first approach, computerized literature searches are

conducted of databases such as CINAHL, MEDLINE, PsycLIT, and SocioFile. The second approach involves scanning nursing, medical, psychological, sociological, dissertation, and other abstracts; dissertation abstracts are particularly useful for locating unpublished studies. In the descendancy approach, citation indexes are used to identify studies that cite papers central to a topic. The ancestry approach involves checking reference lists to identify relevant publications that have not been previously identified. Finally, the invisible college approach is an informal approach, to obtain unpublished studies, conference papers, and government, agency, and foundation reports through professional networks (Cooper). Details of the search procedure, including the approaches used, the years searched, and the search terms used, should be reported so others can evaluate the adequacy of the retrieval process.

Once the search is completed, two investigators independently review all retrieved studies to determine whether they meet the inclusion criteria. Cohen's (1960) kappa correlational statistic is used to determine the extent of agreement between the two investigators; a kappa of .80 is generally considered acceptable (Waltz, Strickland, & Lenz, 1991).

Classifying and Coding Study Characteristics

The third step in a meta-analysis consists of classifying and coding the characteristics of all studies meeting the inclusion criteria. These study characteristics are the independent variables in the meta-analysis and are commonly classified into methodological and substantive features of the studies (Glass, McGaw, & Smith, 1981). Methodological features refers to variables related to the research design and methodology for each study, as well as publication information. Examples of methodological features include research design, sample size, sampling method, attrition rate, degree of blinding of experimenter, rating of study quality, source of the publication (e.g., journal, dissertation, book), date of publication, and form of study (e.g., published or unpublished). Substantive features refers to the research domain or question that the meta-analysis is addressing. Examples include demographic characteristics of the sample in the primary study (e.g., age, gender, ethnicity, health status), theoretical framework for the primary study, type of nursing intervention administered, and the outcome measure.

The study characteristics are then coded to determine whether the meta-analytic findings differ according to the nature of the primary studies. A coding form is developed to ensure a valid and reliable

process of collecting data (McCain et al., 1986). The coding form, often referred to as a codebook, is a compilation of the computer coding information of all the study characteristics. The coding form is valid to the extent that all relevant study characteristics are included, and it is reliable to the extent that the codes are used consistently and accurately. Investigators must be trained to use the coding form, and both inter- and intra-rater reliability should be assessed. All procedures used for ensuring reliability and validity should be as rigorous as those used in primary research and should be reported in detail in the final report (McCain et al.).

Converting the Outcome Measure to a Common Scale or Metric

Pooling the results from individual studies is simplified when research studies measure outcomes with the same instrument. Outcomes can be added up and then divided by the total number of studies to obtain an average effect size. More commonly, though, different investigators study the same construct using instruments that yield numbers on completely different scales. Consequently, a common scale or metric is needed to aggregate findings across studies. Indices of effect magnitude provide this common metric because they do not depend on the arbitrary scaling of the outcome measure. Two scale-free metrics are recommended for analyzing outcomes measured on a continuous scale: the effect size statistic *(d)* and the correlational statistic *(r)*. The discussion that follows will focus on using the effect size *(d)* as an index of effect magnitude. Interested readers should refer to Rosenthal (1984) for a discussion of meta-analysis using the correlational statistic *(r)*.

Effect size *(d)* is a measure of the mean difference between the experimental and the control group measured in standard deviation units (Cohen, 1988). This statistic provides information about the direction and magnitude of the effect, and is used for expressing the effectiveness of experimental treatments. It is interpreted as the number of standard deviation units by which the control group could have benefited or failed to benefit (depending on the sign) had they been exposed to the experimental treatment. When calculating the effect size, it is important to remember that some outcome measures assign a high score for the desired outcome of a treatment, while others assign a low score for the same outcome. It must be ensured that all outcomes in the desired direction have the same sign; otherwise, combining effect sizes will be meaningless. The effect size can be calculated in absolute terms and assigned a positive sign when the experimental group had higher values than the control group (Devine, 1990).

Although the *d* statistic can simply be calculated as the difference between the mean of the treatment group and the mean of the control group divided by the standard deviation, effect sizes that have been adjusted for sample-size discrepancies provide more stable estimates of population effect sizes. The following formula estimates an effect size that is adjusted for sample-size discrepancies (Hedges & Olkin, 1985):

$$d_i = \left(\frac{3}{1 - 4\,(n_e + n_c) - 9}\right)\left(\frac{\bar{x}_e - \bar{x}_e}{Sp}\right)$$

where: d_i = effect size for each individual study
n_e = sample size for the experimental group
n_c = sample size for the control group
$\bar{x}_e$ = mean for the experimental group
$\bar{x}_c$ = mean for the control group
Sp = pooled within-group standard deviation

The effect size is generally standardized by using the pooled within-group standard deviation, since this provides an unbiased estimator of effect size (Hedges & Olkin; Hunter & Schmidt, 1990). However, when the assumption of equal variance between the groups is not satisfied, Glass and his colleagues (1981) recommend using the control-group standard deviation. When means and standard deviations are not available, the effect size can be estimated from other statistics, such as the *t*, F, or correlational statistic (Hedges & Olkin).

Cohen (1988) provides standards for interpreting effect sizes in terms of absolute numbers, but warns that these standards are still being refined. An effect size from .20 to .49 is considered a small effect, .50 to .79 a medium effect, and .80 or greater a large effect.

Studies often contain more than one outcome measure, especially in non-experimental research; however, meta-analytic procedures require the calculation of a single effect size per eligible study (Petitti, 1994). Inclusion of more than one effect size per study inappropriately weights studies and violates the statistical assumption of independence. Several options are available for dealing with multiple outcomes. The researcher may choose the most conceptually congruent outcome, select an outcome randomly, average the effect sizes to provide a single estimate, or use multivariate meta-analytic procedures (Hedges & Olkin, 1985). Whatever option is chosen, it should be decided a priori, used consistently throughout the study, and documented in the study protocol.

The effect size is the dependent variable in the meta-analysis. Despite the fact that common statistics such as means and standard

deviations are used to calculate the effect size, studies with missing data are common (Devine & Cook, 1983; McCain & Lynn, 1990). When the publication includes the author's address, it may be possible to obtain the required data. If the information is unavailable, the study cannot contribute to the summary estimate. However, the study should not be considered ineligible for inclusion in the meta-analysis. Rather, it should be reported as having missing data. Otherwise, researchers evaluating the retrieval process will assume that the study was missed in the literature search (Petitti, 1994).

Aggregating the Study Findings

The next step in the meta-analysis is aggregation of the study findings. Statistical analyses for combining data from several studies involves (1) estimating a summary measure of effect size, (2) estimating the variance of the summary measure, (3) testing for homogeneity, and (4) placing a conference interval around the summary measure.

Estimating a summary measure of effect size. The summary measure of effect size is a descriptive statistic of central tendency providing a single summary value for the effect of an independent variable on a dependent variable within an entire area of study. Of the several methods for estimating this summary statistic, the simplest is to add up all the effect sizes from each individual study and divide the figure by the total number of studies. This provides an unweighted summary measure of effect size. A second method is to weight effect sizes before calculating the summary measure. This is done because Hedges and Olkin (1985) found that d is a slightly biased estimator of effect size. To provide an unbiased estimator, they recommend weighting each effect size by the reciprocal of the estimated variance of d in each of the studies to be aggregated in the meta-analysis. When sample sizes in the experimental and control groups are almost equal and greater than 10, the following formula provided by Rosenthal and Rubin (1982) can be used to estimate the weight of each study:

$$w_i = \frac{2N_i}{8 + d_i^2}$$

where: w_i = weight for each individual study
N_i = total sample size for each individual study
d_i = effect size for each individual study

A weighted summary measure of effect size is then calculated as the sum of the products of the weights times the effect sizes from each

of the individual studies divided by the sum of the weights from each of the individual studies.

$$d_s = \frac{\Sigma(w_i \times d_i)}{\Sigma w_i}$$

where: d_s = weighted summary measure of effect size
w_i = weight for each individual study
d_i = effect size for each individual study

This formula now provides a summary measure that is weighted for sample-size discrepancies and has the smallest possible variance. Frequently, both weighted and unweighed summary measures are reported.

Individual effect sizes that are extreme in relation to the rest of the values (i.e., outliers) distort the summary measure and the observed variance (Huffcutt & Arthur, 1995). One method for dealing with outliers is to use the median effect size, since this measure of central tendency is less sensitive to outliers (Light, 1980). Alternatively, outliers can be "winsorized" – which is a procedure for trimming the data by discarding outliers from both tails of the distribution to make the data set more representative of the population. Light recommends deleting the largest 5% and the smallest 5% of the values and using the remaining effect sizes to calculate the summary measure. In addition, the study characteristics of studies with outlying effect sizes should always be examined to determine why they are atypical. These studies may reveal interesting patterns that could contribute significant information about an intervention. For example, it may be that an intervention was more effective for certain subgroups of the population or in certain settings.

Variance of the summary measure. The variance of the summary measure is calculated to provide an index of the variability associated with the summary measure. If the variance is large, the summary effect size may be misleading when interpreted in isolation. A large variance suggests the presence of confounding variables. The variance is estimated by calculating the inverse of the sum of all the weights for each independent study, as follows (Rosenthal & Rubin, 1982):

$$\sigma_s^2 = \frac{1}{\Sigma w_i}$$

where: σ_s^2 = variance of the summary measure of effect size
w_i = weight for each individual study

Testing for homogeneity. Although the summary measure of effect size provides important descriptive information, it can be interpreted with confidence only if the effect sizes from the individual studies are homogeneous. Testing for homogeneity is essentially an attempt to discover whether the variation in effect sizes can be attributed to sampling error. This tests the hypothesis that all effect sizes are equal against the alternative hypothesis that at least one effect size is different (Hedges & Olkin, 1985). It is computed using a Q statistic, which has an asymptotic chi-square distribution with degrees of freedom equal to the number of studies minus one (df = k - 1), and is calculated as sum of the weights multiplied by the squared difference between the summary effect size and the individual effect sizes:

$$Q = \Sigma\,[w_i(d_i - d_s)^2]$$

where: d_i = effect size for each individual study
d_s = summary measure of effect size
w_i = weight for each individual study

A nonsignificant Q statistic indicates that the effect sizes are homogeneous (i.e., variation is due to sampling error) and that effect sizes can be pooled, since they estimate the same population parameter. A statistically significant Q statistic, however, indicates rejection of the null hypothesis that the study effect sizes are homogeneous. Since the effect sizes are heterogeneous, they do not estimate the same population parameter and should not be pooled.

Once again, efforts should be made to search for the study characteristics that account for the variability. It may be that the primary studies were not testing the same hypothesis and should not be included in the same meta-analysis. Alternatively, some particular study characteristic (e.g., subject gender, length of treatment, or type of study) may account for the variability. When certain study characteristics mediate the relationship between the treatment and the outcome, consideration should be given to how these mediating variables might explain the heterogeneity. When heterogeneity exists, Hedges and Olkin (1985) advocate clustering the effect sizes into more homogeneous groups, and testing for homogeneity within these subgroups.

It should be noted, however, that testing for homogeneity is a controversial technique in meta-analysis (Hunter & Schmidt, 1990; Petitti, 1994). Hunter and Schmidt criticize this technique because it focuses solely on sampling error and ignores other artifactual sources of between-study variation such as that caused by computational or tran-

scriptional errors, or differences between studies in reliability of measurement. Further, they caution that even when effect sizes are the same across studies, artifactual sources of variation alone may create variance beyond sampling error, causing a significant test when the sample size is large and statistical power is high. For further discussion of the controversy and the various approaches that can be taken when heterogeneity exists, refer to Hunter and Schmidt and to Petitti.

Confidence intervals. Although the summary measure is the "best" estimate of the true effect size, confidence intervals provide an estimate of the possible range of values for effect sizes within a given probability (Hedges, 1982). By convention, 95% confidence intervals are generally used. A confidence interval that does not include zero indicates that the summary measure is significantly different from zero. As the width of the confidence interval increases, less confidence can be placed in the summary measure. A 95% confidence interval (CI) for the summary measure of effect size is calculated as follows:

$$95\%CI = d_s \pm (1.96 \sqrt{\sigma_s^2})$$

where: d_s = summary measure of effect size
σ_s^2 = variance of the summary measure of effect size

Interpreting the Study Findings

The final step in the meta-analytic procedure is relating the findings to the study characteristics in order to explain the results theoretically and to discuss its implications. Although meta-analysis provides a quantitative approach to resolving contradictory findings, the investigator must have a thorough understanding of the substantive area to make sense of the results and to derive meaningful conclusions. The quantitative findings of the meta-analysis must be discussed in relation to the current level of knowledge. The final report should include a discussion of the results, including the implications, and recommendations for further primary research or additional meta-analyses (Smith, 1994).

Advantages of Meta-Analysis

When rigorously used, meta-analysis can lead to an improved quality of research integration. It is more systematic, explicit, and objective than other methods currently used to summarize data (Mintz, 1983; Rosenthal, 1984). This approach uses formal procedures for combining the findings from empirical studies and requires fewer subjective judge-

ments. Because the study protocol is explicitly documented in the final report, other researchers can evaluate the adequacy of the method – which is important if the findings are to establish evidence of the efficacy of interventions. In addition, the explicit nature of the procedure allows for replication, a significant aspect of scientific inquiry.

A meta-analysis is more likely than other methods to lead to summary statements of greater thoroughness and precision. Meta-analytic results provide a more precise statement about the magnitude of effectiveness of an intervention, sample variability, and the interrelationships between variables and differences, as well as lists of descriptive data that allow for the identification of patterns (Smith, 1994). While summary measures of effect size provide important information about the effectiveness of interventions, the variation in effectiveness may also help identify particular groups that did or did not benefit from the experimental condition, thus leading to insightful information and areas for further research.

Conclusions reached in meta-analytic reviews are more definitive about the effects of an intervention than are those of narrative reviews (Devine, 1990). Because meta-analytic studies have more statistical power than primary studies, they are more likely to detect a consistent treatment effect, even when the power in the primary studies is low (Devine; Petitti, 1994). This feature is a particularly important one for clinical nursing research. It is often difficult to obtain large enough samples and sufficient statistical power to detect clinically relevant effects in a single study (Devine). The greater power of meta-analysis means less likelihood of accepting that there is no effect or relationship when an intervention really is effective (i.e., less chance of Type II errors).

Finally, a meta-analysis provides a more thorough description of the current status of research in a domain, identifies gaps in the knowledge base (Fiske, 1983), and gives directions for further research, by generating hypotheses for more primary research or additional meta-analyses (Smith, 1994). This method can help resolve uncertainty about the effects of interventions, which is essential given the current emphasis on establishing evidenced-based practice. When studies do not fit together well, researchers can better determine, when using the meta-analytic approach, where the inconsistencies and incongruencies lie. When a few studies present findings that are in marked contrast to the rest, isolating them may result in the identification of certain common characteristics, leading to discovery of meaningful information for designing more effective interventions. A major strength of this tech-

nique is that it encourages the researcher to view conflicting findings constructively (Light, 1980; Pillemer & Light, 1980).

Methodological Issues

Meta-analytic techniques are still relatively new and are continually being refined. The techniques presented above were based on Glass and his colleagues (1981), Hedges and Olkin (1985), and Rosenthal (1984), but other methodologies considered rigorous have been developed as well (Hunter & Schmidt, 1990). Selection of an approach is influenced by the research question, the level of data, and the educational background of the researcher. Regardless of the meta-analytic approach used, certain methodological issues are common to all approaches, and a researcher should be aware of these when planning to conduct a meta-analysis or when interpreting the findings of a meta-analysis.

Adequacy of the Database

A primary concern in meta-analysis is that the database be representative of all studies conducted on a phenomenon of interest. External validity is threatened when important strata of the population of studies are missing. Unfortunately, though, representativeness is difficult to ensure. Even when the topic area is well defined, it is often difficult to locate all relevant studies (Oxman & Guyatt, 1988). They may be in press or not yet indexed in computerized databases. Papers presented at conferences, theses, dissertations, government studies, and other unpublished studies are difficult to identify and obtain. Because of its limited circulation, such literature is often referred to as fugitive literature (Smith, 1980). Finally, in most disciplines there is a tendency to publish only research with significant findings. Rosenthal (1980) describes this phenomenon as the "file drawer problem," because studies that fail to reach statistical significance are more likely to remain in researcher's file drawers. When a meta-analysis uses only published studies, the sample may be biased; results of the meta-analysis will be skewed toward positive findings (Hunter & Schmidt, 1990; Lynn, 1989). To minimize this bias, every effort should be made to retrieve unpublished reports. When only published studies are used, this limitation of the meta-analysis should be acknowledged.

Variation in Study Quality

Another important issue in meta-analysis concerns variation in the quality of studies incorporated into the analysis. There is ongoing

debate whether all studies on a given topic should be included or whether inclusion should depend on a certain level of study quality. Glass (1976) argues that, to avoid a systematic investigator bias, all studies should be included. Other meta-analysts argue that there should be some differentiation of studies based on quality, because meta-analytic procedures are insensitive to the validity of the findings of primary research (Brown, 1991; Hedges & Olkin, 1985; Petitti, 1994; Rosenthal, 1984). Studies with low validity will contribute just as much to the summary measure as studies with comparable sample size but greater validity. Further, including studies of poorer quality may yield information that is not valid, which threatens the validity of the meta-analytic findings.

There is strong potential in meta-analysis for uncontrolled validity threats if many of the aggregated studies were poor in quality for the same reason (Petitti, 1994). For example, when several studies failed to randomly assign subjects to treatment groups, there is greater chance for error that is non-random and systematic. Statistical aggregation methods cannot overcome problems of bias and uncontrolled confounding. This is potentially a serious problem in any meta-analysis involving non-experimental studies that lack randomization to minimize bias and confounding.

One strategy to deal with this issue is to assess and code study quality as an additional study characteristic, and to examine the relationship between study quality and effect size (Mintz, 1983; Petitti, 1994; Rosenthal, 1984). The decision to include or exclude studies of poorer quality may be facilitated by examining the correlation between effect sizes and ratings of study quality. If the rating of study quality does not correlate with the effect size, there is less reason to exclude the study from the analysis. If a correlation is observed, the conclusions of higher-quality studies should be given greater weight, since these studies are more likely to yield valid information (Mintz). A system of weighting studies based on their rating of study quality can be employed (Rosenthal, 1984). Studies of low quality will be given a weight of zero and contribute no information. An alternative strategy is to stratify effect sizes based on the rating of study quality, and then examine effect sizes within each stratum.

Several instruments have been developed to rate research quality (Chalmers et al., 1981; Duffy, 1985; Smith, 1988). Research quality can also be rated using Cook and Campbell's (1979) list of threats to internal and external validity and coding studies for the presence or absence of these threats (Cooper, 1982; Mitchell, 1985). Rating study quality

often needs to be specifically tailored for the type of research the meta-analysis is addressing. Criteria that define quality have to be identified and a standardized system of rating each criterion has to be devised. To minimize investigator bias, two raters who are blinded to the study investigator, the affiliated institution, and the journal of publication should independently rate study quality according to the predefined criteria (Rosenthal, 1984). Inter- and intra-rater reliability should be assessed to ensure consistency in ratings.

Variation in Method

A common criticism of meta-analysis is that the procedure aggregates data from studies in which the independent variable, the dependent variable, and the sampling units are not uniform (Rosenthal, 1984). For example, in Blegen's (1993) meta-analysis of 48 studies to determine factors contributing to job satisfaction among nurses, satisfaction was measured using 21 different instruments. In Brown and Hellings's (1988) meta-analysis of 10 studies that examined early maternal-infant contact and attachment behaviour, the dependent variable varied considerably from observations of maternal gazing, affectionate behaviour, tender touching, interaction and stimulation, demonstrating proximal behaviours, and using a mother-infant feeding profile. When the conceptual congruence of the dependent variables used in the primary studies are debatable, aggregating the effect sizes to provide a single summary measure of effectiveness may be theoretically meaningless and will not provide sound evidence for establishing an evidential base for nursing practice. Concern about conceptual congruence – an important issue in any discipline – may have special relevance for nursing research, because constructs from other disciplines are frequently used. Even constructs with the same name may differ significantly in conceptual meaning across disciplines (Smith, 1994).

Concern about conceptual and methodological variation is commonly referred to as the "apples and oranges" issue. Glass and his colleagues (1981) argue that it is no more problematic to pool data across studies than it is to generalize across subjects in primary research, and that if interactions between study characteristics and effect sizes are suspected, they can be tested in the context of a meta-analysis. Other investigators (Moody, 1990; Slavin, 1984) argue that it is inappropriate to combine studies when the settings, subjects, or empirical qualities are drastically different. For a meta-analysis to be valid, the studies combined must address common hypotheses and be conceptually and methodologically equivalent. Further, when the primary studies are

similar, then differences in effect sizes are more likely to be explained by chance (Oxman & Guyatt, 1988).

Conclusion

With the increasing amount of research being conducted in nursing, integrating findings systematically is integral to understanding the current status of knowledge in an area and to obtaining direction for further investigative efforts. Meta-analysis provides for rigorous synthesizing and integrating the results of a large body of literature and determining the efficacy of nursing interventions, and is also a means for incorporating the information obtained from individual studies into the discipline's knowledge base.

While meta-analysis represents a significant improvement over traditional methods of reviewing research, it is limited by the methodological and theoretical constraints of the primary studies included in the analysis. Nevertheless, when meta-analytic techniques are rigorously applied, they provide researchers with a powerful tool to help make sense of nursing's increasing body of research.

References

Blegen, M.A. (1993). Nurses' job satisfaction: A meta-analysis of related variables. *Nursing Research, 42*(1), 36-41.

Brown, S. (1991). Measurement of quality of primary studies for meta-analysis. *Nursing Research, 40*(6), 352-355.

Brown, S., & Hellings, P. (1988). A case study of qualitative versus quantitative reviews: The maternal-infant bonding controversy. *Journal of Pediatric Nursing, 4*(2), 104-111.

Chalmers, T.C., Smith, H. Jr., Blackburn, B., Silverman, B., Schroeder, B., Reitman, D., & Ambroz, A. (1981). A method for assessing the quality of a randomized control trial. *Controlled Clinical Trials, 2*, 31-49.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46.

Cohen, J. (1988). *Statistical power for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cook, T.D., & Campbell, D.T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.

Cooper, H.M. (1982). Scientific guidelines for conducting integrative research reviews. *Review of Educational Research, 52*(2), 291-302.

Curlette, W.L., & Cannella, K.S. (1985). Going beyond the narrative summarization of research findings: A meta-analysis approach. *Research in Nursing & Health, 8*, 293-301.

Devine, E.C. (1990). Meta-analysis: A new approach for reviewing research. In N.L. Chaska (Ed.), *The nursing profession: Turning points* (pp. 180-185). St. Louis: Mosby.

Devine, E.C., & Cook, T.D. (1983). A meta-analytic analysis of effects of psychoeducational interventions on length of postsurgical hospital stay. *Nursing Research, 32*(5), 267-274.

Dickerson, K. (1990). The existence of publication bias and risk factors for its occurrence. *Journal of the American Medical Association, 263*(10), 1385-1389.

Duffy, M.E. (1985). A research appraisal checklist for evaluating nursing research reports. *Nursing & Health Care, 6*, 539-547.

Fiske, D.W. (1983). The meta-analytic revolution in outcome research. *Journal of Consulting and Clinical Psychology, 51*(1), 65-70.

Glass, G. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*, 3-8.

Glass, G., McGaw, B., & Smith, M.L. (1981). *Meta-analysis in social research.* Beverly Hills, CA: Sage.

Hedges, L.V. (1982). Estimation of effect size from a series of independent experiments. *Psychological Bulletin, 92*(2), 490-499.

Hedges, L.V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Toronto: Academic Press.

Huffcutt, A.I., & Arthur, W. (1995). Development of a new outlier statistic for meta-analytic data. *Journal of Applied Psychology, 80*(2), 327-334.

Hunter, J.E., & Schmidt, F.L. (1990). *Methods of meta-analysis.* Newbury Park, CA: Sage.

Light, R.J. (1980). Synthesis methods: Some judgement calls that must be made. *Evaluation in Education, 4*, 13-17.

Lynn, M.R. (1989). Meta-analysis: Appropriate tool for the integration of nursing research. *Nursing Research, 38*(5), 302-305.

McCain, N.L., & Lynn, M.R. (1990). Meta-analysis of a narrative review: Studies evaluating patient teaching. *Western Journal of Nursing Research, 12*(3), 347-358.

McCain, N.L., Smith, M.C., & Abraham, I.L. (1986). Meta-analyses of nursing interventions: The codebook as a research instrument. *Western Journal of Nursing Research, 8*(2), 155-167.

Mintz, J. (1983). Integrating research evidence: A commentary on meta-analysis. *Journal of Consulting and Clinical Psychology, 51*(1), 71-75.

Mitchell, T.R. (1985). An evaluation of the validity of correlational research conducted in organizations. *Academy of Management Review, 10*(2), 192-205.

Moody, L.E. (1990). Meta-analysis: Qualitative and quantitative methods. In L.E. Moody (Ed.), *Advancing nursing science through research,* Vol. 2 (pp. 70-110). London: Sage.

Oxman, A.D., & Guyatt, G.H. (1988). Guidelines for reading literature reviews. *Canadian Medical Association Journal, 138*, 697-703.

Petitti, D.B. (1994). *Meta-analysis, decision analysis, and cost-effectiveness analysis.* New York: Oxford University Press.

Pillemer, D.B., & Light, R.J. (1980). Synthesizing outcomes: How to use research evidence from many studies. *Harvard Educational Review, 50*(2), 176-195.

Reynolds, N.R., Timmerman, G., Anderson, J., & Stevenson, J.S. (1992). Meta-analysis for descriptive research. *Research in Nursing & Health, 15,* 467-475.

Rosenthal, R. (1980). Combining probabilities and the file drawer problem. *Evaluation in Education, 4,* 18-21.

Rosenthal, R. (1984). *Meta-analytic procedures for social research: Applied social research methods series.* London: Sage.

Rosenthal, R., & Rubin, D.B. (1982). Comparing effect sizes of independent studies. *Psychological Bulletin, 92,* 500-504.

Slavin, R.E. (1984). Meta-analysis in education: How has it been used? *Educational Researcher, 13*(18), 6-15.

Smith, M. (1980). Publication bias and meta-analysis. *Evaluation in Education, 4,* 22-24.

Smith, M. (1988). *Meta-analysis of nursing intervention research.* Birmingham, AL: Author.

Smith, M. (1994). Meta-analysis. In B. Sarter (Ed.), *Paths to knowledge: Innovative research methods for nursing* (pp. 77-91). New York: National League for Nursing.

Smith, M.C., & Naftel, D.C. (1984). Meta-analysis: A perspective for research synthesis. *Image, 16*(1), 9-13.

Waltz, C.F., Strickland, O.L., & Lenz, E.R. (1991). *Measurement in nursing research* (2nd ed.). Philadelphia: F.A. Davis.

Acknowledgements

This work is supported by a Health Research Studentship from the Alberta Heritage Foundation for Medical Research (AHFMR). The author thanks Marilyn Hodgins, Carole Estabrooks, Dr. Louise Jensen, and Dr. Steve Hunka for reviewing and commenting on this article in its formative stages.

Date accepted: September 1996