Exploratory meta-analysis or research synthesis has been advocated as a way of developing important hypotheses for further study. An exploratory research synthesis was conducted on the carotid endarterectomy (CE) literature to illustrate this method. The CE scientific literature is similar to that of many other new medical interventions because it contains numerous limitations to data quality. Exploratory research synthesis of such literature necessitates a number of methodological and statistical considerations to address these limitations, including the problems of missing data, appropriate unit of analysis, nonnormal distribution of outcomes, and lack of controlled studies. Strengths and limitations of the exploratory research synthesis approach are discussed within the context of public policy decisions for assessing medical technologies.

EXPLORATORY RESEARCH SYNTHESIS

Methodological Considerations for Addressing Limitations in Data Quality

WILLIAM H. YEATON
University of Michigan
JOHN C. LANGENBRUNNER
U.S. Office of Management and Budget
JOSHUA M. SMYTH
PAUL M. WORTMAN
State University of New York at Stony Brook

AUTHORS' NOTE: The authors thank Ingram Olkin and Anant Kshirsagar for the many insightful suggestions on the statistical techniques to use in conducting this synthesis; Daion Smith for her assistance with the data collection; and Barker Bausell and an anonymous reviewer. The views expressed in this paper are strictly those of the authors and do not necessarily represent the positions of organizations with which they are affiliated. Preparation of this article was supported in part by Grant No. HS06264 to Paul Wortman from the Agency for Health Care Policy and Research. Request for reprints should be sent to Paul Wortman, Department of Psychology, SUNY-Stony Brook, Stony Brook, NY 11794-2500.

EVALUATION & THE HEALTH PROFESSIONS, Vol. 18 No. 3, September 1995 283-303

The past decade has seen the development, application, and widespread adoption of statistical techniques for aggregating the results from many research studies (Hedges & Olkin, 1985). This method generally has been termed *meta-analysis* (Glass, 1976) or *research synthesis* (Cooper & Hedges, 1994; Wortman & Yeaton, 1987). The diffusion of the research synthesis approach has been rapid (Glass, McGaw, & Smith, 1981; Hedges & Olkin, 1985; Hunter & Schmidt, 1990); within a decade it has spread across many disciplinary areas, including health and the evaluation of medical technology (Louis, Fineberg, & Mosteller, 1985; Sacks, Berrier, Reitman, Ancona-Berk, & Chalmers, 1987).

One form of synthesis, exploratory research synthesis (Wortman, 1994), has also been advocated as an approach to better link medical research with decision making and policy development (e.g., ISIS-2 Collaborative Group, 1988, p. 357). As medical procedures and technologies have assumed an increasingly central role in the delivery and costs of health services over the last 2 decades (Office of Technology Assessment, 1984), researchers and policymakers have recognized the need for less costly techniques that draw upon existing scientific knowledge and clinical understanding (Fuchs & Garber, 1990; Roper, Winkenwerder, Hackbarth, & Krakauer, 1988).

For example, information regarding coverage and reimbursement of procedures by third-party payers has been typically derived from a traditional literature review of research studies that vary in methodologic adequacy and applicability for decision making (Office of Technology Assessment, 1982, 1984). In this context, the government of the United States has called for a more widespread application of such synthesis approaches as meta-analysis to aggregate the results of numerous scientific studies (Office of Technology Assessment, 1982, p. 17). A major obstacle to conducting such syntheses is that the scientific literature for most emerging technologies consists entirely of weak preexperimental or quasi-experimental designs (Campbell & Stanley, 1966).

The purpose of this article is to examine the methodological considerations in conducting an exploratory research synthesis of poorly controlled studies. It examines the validity and utility of the exploratory synthesis approach as a viable meta-analytic variant for such

preexperimental and quasi-experimental data. A number of methodological problems were encountered in the course of conducting the synthesis, and analytic approaches were developed to maximize the validity of the findings (for specific results of this synthesis of the carotid endarterectomy literature, see Langenbrunner, Wortman, Yeaton, Kshirsagar, & Holloway, 1993).

An exploratory research synthesis of carotid endarterectomy (CE) will be used to illustrate the methodology. CE, first established in the mid-1950s, has become a relatively high-volume, but controversial, surgical operation (e.g., Agency for Health Care and Policy Research [AHCPR], 1991; Callow et al., 1988). The procedure attempts to enlarge the interior space of the carotid artery and to remove plaque from the arterial wall. An incision is made in the neck to expose the narrowed portion of the artery. In some instances, a shunting procedure is used in which tubing carries blood around the area of the surgery, and the diseased segment is opened, cleaned, and sutured together. Blood flow through the repaired artery is then restored (Kistler, 1989). Traditionally, the operation has been performed on patients with signs or symptoms of cerebrovascular disease in an attempt to prevent stroke and ischemic attack (Merrick et al., 1987).

The CE synthesis was initiated within an important policy context and guided by the assistance of physicians working to develop Medicare coverage recommendations. The scientific literature consisted largely of reports of uncontrolled, retrospective, surgical series. The only completed, well-designed study was the North American Symptomatic Carotid Endarterectomy Trial (1991) or NASCET. It reported a statistically significant surgical benefit after 2 years in symptomatic patients with 70%-99% carotid stenosis. However, the fastest growth in CE among Medicare recipients has been in asymptomatic patients and patients with moderate stenosis. Consequently, this exploratory research synthesis focused on asymptomatic patients.

METHOD: OBJECTIVES AND LIMITATIONS

Research synthesis is considered to be inferentially strongest when confined to randomized studies (Sacks et al., 1987). However, even

randomized studies are not immune to serious validity problems, including differential patient attrition and the diffusion of the treatment to control group patients (Wortman, 1994). More recently, synthesis techniques have been used with both nonrandomized, prospective studies (Ozminkowski, Wortman, & Roloff, 1988; Wortman & Yeaton, 1987) and retrospective, case control designs (Longnecker, Berlin, Orza, & Chalmers, 1988). In fact, there now exists a recognition of the need to combine the results of studies using research designs of varying quality (U.S. General Accounting Office, 1992). A number of limitations were recognized in the course of developing the CE database and conducting the synthesis. They are discussed below. This study assesses the viability of extending synthesis methods to emerging medical technologies like CE where the scientific research literature is characterized by relatively weak research designs.

DEVELOPING THE CE DATABASE

A sizable literature of 214 studies published between 1968 and 1988 was identified. Only 19 of these studies used control groups, including three randomized studies. Of the 195 uncontrolled studies, 128 were conducted with patients who were consecutively enrolled. A total of 364 study-level groups and subgroups of patients with similar indications for CE were identified.

An article was included in the database if it was written in English and reported original data with at least one morbidity or mortality outcome measure. Studies reporting CE patients concurrently receiving other surgical interventions (e.g., coronary artery bypass graft surgery) were excluded to allow inference to the CE procedure alone. All relevant CE reports and studies published from 1960 through 1988 were identified from a combination of MEDLINE searches, reviews of the literature, and the reference sections of articles included in the synthesis.

A coding sheet was developed to identify and define variables of interest and importance and to extract pertinent information on study design and methods, treatment intervention, patient characteristics, clinical presentations, and study outcomes during different follow-up periods. Numerical or categorical data for 185 variables were ex-

tracted from the 214 studies. For example, a study was coded as controlled if it used a randomized, cohort, case control or other design involving a no-treatment group. Outcomes were also coded for seven clinical indications using definitions from the RAND consensus study of appropriateness (Merrick et al., 1986)—carotid transient ischemic attack (TIA), carotid mild stroke, stroke-in-evolution, vertebrobasilar TIA, retinal TIA, asymptomatic, and asymptomatic with other surgery planned. From a clinical perspective, these subgroups of outcomes allowed adjustment among study outcomes for important case-mix differences. In addition, aggregate data were obtained by combining results across all patients with similar indications.

A number of data exclusion criteria were used. When multiple reports of the same patients were published, only the most recent findings were included in the database. Data reported in a form that differed from the definitions in the research protocol were coded as "not reported." In addition, aggregate data that could not be separated into categories that fit the definition of variables in the synthesis were discarded from the final working database.

Interrater reliability was assessed by drawing a 10% random sample of studies from the database. Studies were coded independently by three researchers (including a physician and two graduate students) following extensive training on articles representing the full range of coding decisions on all variables in the database. Measures of interrater reliability were computed using an approach developed by Yeaton and Wortman (1993) that takes into account dependencies among coded variables. The overall average agreement among coders for the studies was 86%, with a range of 92% to 78% agreement for variables with differing degrees of dependence.

LIMITATIONS OF PREEXPERIMENTAL DATA

The preexperimental nature (Campbell & Stanley, 1966) of the CE literature imposed several methodologic limitations. The relative absence of controlled studies meant that computation of effect size was not possible. Thus no meta-analytic measures of efficacy or effectiveness were generated, since these involve differences between treated and untreated groups (Hedges & Olkin, 1985). However, it was

possible to determine proxy measures for effectiveness by comparing mortality rates with norms established by physician experts (Langenbrunner et al., 1993). This approach also allows one to address related questions such as the effect on CE mortality of various research circumstances (e.g., type of hospital, patient indication, etc.), and patient populations (e.g., age group).

Descriptive statistical approaches that aggregated outcome rates and proportions were used, and multivariate techniques were used to describe the relationship between selected independent and dependent variables (e.g., patient age and mortality). In performing these analyses across clinical indications, variables such as patient characteristics were coded at both the study level *and* the subgroup level, if necessary. This meant that any statistical correlation of study level predictor variables, such as age, with outcomes at the subgroup level, such as specific indication (e.g., TIA), assumed a similar distribution of values for cases at both levels (if age data were not reported within both the study and the subgroup). For example, the relationship between age (reported at the study level) and mortality for asymptomatic patients (reported at the subgroup level) assumed that the average age in both the study and subgroup levels was identical.

GENERAL ANALYTIC ISSUES

The uneven quality of the information reported in published studies is one of the major problems in conducting a research synthesis (Orwin & Cordray, 1985). Three general methodologic areas were identified as critical to draw valid inference from the CE synthesis: missing data, statistical issues, and approaches for explaining variability of outcomes. The major purpose of this article is to describe and evaluate a number of methodological approaches to these problems. The first area dealt with issues surrounding missing data.

MISSING DATA

Often, data concerning important clinical and outcome variables were missing from studies. In the context of outcome-based, statistical

analyses, such missing data are especially important. Table 1 displays the percentage with which different categories of variables (in Column 1) were reported, as well as their definition or description (in Column 5). Results from Table 1 reveal that most CE information was infrequently reported. Except for the pre-op technique category (i.e., the procedure for examining the arteries), the percentage of variables reported per study is low across all categories, including outcomes, and the overall percentage reported is also quite low at 29%. Moreover, there is little apparent reporting difference by study design. For the 19 controlled studies (see Column 4 of Table 1), reporting quality is higher for some categories, while for others it is less complete.

Table 2 presents percentages of variables reported in the 214 studies and an index made up of a subset of the 22 most important variables used in previous CE analyses from the perspective of Medicare policy formulation (Langenbrunner et al., 1993). These 22 variables were chosen from all the variable categories listed in Table 1. These policy-relevant variables include such measures as mean patient age, gender mix, provider setting (e.g., university/teaching hospital vs. community hospital), 30-day postsurgical mortality and selected 30-day morbidity measures. This index of "percentage reported" was aggregated over all studies by publication year to examine whether reporting practices have improved over time.

Not surprisingly, the percentages of these more policy-relevant measures are consistently higher than those found in Table 1, and consistently higher than the percentage reported number for the "all other variables" category in Table 2. At the same time, Table 2 shows that reporting quality did not improve appreciably over time, either for the index of 22 variables or for all CE database variables. Even in recent years when the number of published articles greatly increased, reporting quality remained virtually unchanged.

Overall, these results are consistent with previous reports (Mosteller, Gilbert, & McPeek, 1980) that have identified similar problems in terms of both clinical and methodological information reported in clinical trials. The reporting quality of these CE studies is also similar to findings by Venulet, Blattner, Bulow, and Berneker (1982), who indicated that only 19% of studies contained a minimum set of information to assess adverse drug reactions.

TABLE 1
Reporting Quality for Carotid Endarterectomy Studies by Types of Variables

Category	Number of Variables	Percent Reported: All Studies (N = 214)	Percent Reported: Controlled (n = 19)	Description of Variables
Pre-op technique	1	76	89	Angiography, Doppler, etc.
Surgical technique	13	35	14	Type of anesthesia, use of shunt, etc.
Patient characteristics	11	40	36	Number of patients, operations, age, etc.
Vessel disease	16	14	9	Ulcerative lesions, site, stenosis, etc.
Outcomes				
Perioperative	21	12	8	Mortality and morbidity (e.g., TIA, MI, etc.)
30 day	21	21	20	
> 30 day	19	11	20	
Overall	185	29	27	

TABLE 2
Reporting Quality for Carotid Endarterectomy Studies by Year of Study Publication

Time Period	Number of Studies Published	Percentage Reported: Index of Variables (n = 22)	Percentage Reported: All Other Variables (n = 163)
Early			
1968-1976	8	43	29
Middle			
1977-1982	68	43	25
Late			
1983-1988	138	45	27
Overall	214	44	26

Approaches to Adjust for Missing Data

Several approaches were employed to address analytic concerns related to missing data. The first approach involved estimation based on the relationship between variables representing similar constructs.

The sole example in this category involved estimation of the number of patients in a study on the basis of the reported number of CE operations. Obviously, the number of patients in a study is a crucial variable for computation of mortality rates and morbidity incidence. Yet it was not reported in over 10% of the 214 studies. (In those cases, the number of CE operations was provided.) For some subgroups such as asymptomatics, 9 of 32 subgroups (28%) did not report the number of patients. Therefore, an estimate of the number of unreported patients in a given study was based on the relationship between patient sample size and numbers of CE operations in the entire database. Regression analyses confirmed that sample size could be predicted with great precision from the total number of CE operations performed (R^2 was consistently greater than .95).

Patients can have either one (i.e., unilateral) or two (i.e., bilateral) CE operations. Additional descriptive statistics were generated for studies in which both variables were reported to further establish the estimate of the number of patients from the number of operations performed. Scatter plots (not shown) were generated for selected outcome events (30-day mortality rate and 30-day morbidity incidence of stroke and myocardial infarction) by ratio increments of .1 from 1.0 to 2.0 of the number of operations to the number of patients. The overwhelming majority of these ratios occurred between 1.1 and 1.3, and the ratios failed to indicate a strong relationship with the 30-day mortality rate. Consequently, we used 1.2 as a way to convert number of operations into number of patients (i.e., number of operations divided by 1.2).

A second, more commonly employed approach to data imputation used a mean value estimate. Given the relatively small set of observations for several of the analyses conducted (e.g., 32 studies for the subgroup of asymptomatic CE patients) and variables in these analyses (e.g., average age of the patient in the sample), estimated mean values were used. This was considered appropriate both because the proportion of missing values was small in these subgroups and because the variance for the set of observations was relatively small as well. For example, average age was 69 for both the asymptomatic subgroup of studies and the overall group of 214 studies; variance was only 11.8 for asymptomatics versus 18.2 for the larger overall group. The use of

estimated mean values would reduce measures of variability, while leaving the estimate of the variable mean unchanged.

A third, related approach produced estimates based on the availability of reported values for variables defined in only one way. For example, two variables were specified for use of surgical shunt during the CE operation itself. The first was coded as "yes" or "no" depending on whether or not the use of shunt was reported in each study. The second variable reported the percentage of patients actually receiving the shunt. For analytic purposes, if the first variable was coded, but not the second, a simple rule was followed: if yes, report 100%, if no, report 0%. Sensitivity analysis was undertaken that consisted of an examination of the original distribution of reported values for the second variable as compared with the use of the 100%/ 0% imputation rule. More than half the reported values for this second variable were coded 0% or 100%, thus providing assurance that estimation using this rule for missing values would not systematically alter overall reporting patterns.

The fourth and final approach for addressing missing values was to combine closely related variables into a single hybrid variable. The clearest example of this approach involved outcomes for mortality or selected morbidity measures such as stroke. Outcome variables were initially specified and stratified by follow-up periods—perioperative (within 24 hours), 24 hours to 30-days postsurgery, and beyond 30-days postsurgery. The first two categories were combined as a short-term outcome construct; indeed, CE study authors most often reported either perioperative or 30-day outcomes rather than both. (Care was taken not to double count deaths when mortality was reported in both periods.) Such a hybrid value is preferable to estimation of mean values, especially when a large number of zero values are reported. Reports of a greater than 30-day follow-up period were very inconsistent, thus making any systematic aggregation problematic.

These four approaches were used to address missing data problems in key variables such as mortality (less than 30 days) and use of shunt. In particular, they permitted a critical set of multivariate analyses to be conducted showing the potential importance of the use of a shunt on CE mortality (Langenbrunner et al., 1993). Nevertheless, a substantial number of variables had such a large proportion of missing

values that no statistical approach was adequate, and these variables were excluded from all analyses. Multivariate analyses typically were undertaken with a pool of less than 10 to 12 variables for which values were consistently reported. This number of variables represented less than 5% of the overall number of variables initially specified for data collection and coding. The small number of variables available for multivariate analyses is discussed in more detail in later sections.

STATISTICAL ISSUES

The second general analytic area concerned issues related to two aspects of statistical conclusion validity (Cook & Campbell, 1979). Findings may depend on whether results are aggregated using the study or the individual case as the unit of analysis. When all reported cases in a series are treated by the same physician or in a single hospital, questions are raised about the independence of case results. On the other hand, if all studies are weighted equally regardless of number of cases in each study, statistical power is diminished. The second potential threat to statistical conclusion validity relates to whether values for individual variables were normally distributed, a critical assumption in multivariate analysis that can otherwise result in violated assumptions of statistical tests.

UNIT OF ANALYSIS

The first issue focused on variations in outcomes depending on the unit of analysis or weighting of data used in computations. Hedges and Olkin (1985) discuss the possible bias stemming from choice of the unit of analysis. In the case of CE, two alternative analytic approaches were employed using either (a) the individual patient or (b) the individual study as the unit of analysis. Each approach had its unique strengths and limitations.

By using the individual cases in each study, pooling across studies greatly increased sample size and statistical power. This approach is especially helpful when analyses are focused on subgroups of patients with similar clinical indications. For example, 30-day mortality results

TABLE 3
Mean 30-Day Mortality Rates (standard deviations) for Carotid Endarterectomy

Indication[a]	*Percentage Unweighted*[b]	*Percentage Weighted*[c]
Carotid TIA[d]	1.46 (2.09)	0.90 (0.13)
	(*n* = 19)	(*n* = 6068)
Carotid mild stroke	0 (0)	0 (0)
	(*n* = 1)	(*n* = 34)
Asymptomatic	0.78 (1.24)	1.28* (0.33)
	(*n* = 20)	(*n* = 1250)
Vertebrobasilar TIA	0.20 (0.41)	0.58 (0.58)
	(*n* = 4)	(*n* = 483)

NOTE: The *n*s for unweighted data refer to the number of studies; the *n*s for weighted data refer to the number of reported patient cases of CE.
a. Outcomes were also computed for clinical indication categories of stroke-in-evolution, asymptomatic, other surgery planned, and retinal TIA; sample sizes were too small to be meaningful.
b. Unweighted refers to the use of the individual study as the unit of analysis.
c. Weighted results use the individual case as the unit of analysis.
d. TIA = transient ischemic attack.
*Indicates statistical significance between unweighted and weighted means at the $p < .10$ level.

with CE were reported for 1,250 asymptomatic patients (see Table 3). In the CE database, study samples varied from 3 to 3,574, and the distribution of sample sizes was skewed toward the lower end of this range. Nearly 75% of the studies had samples of less than 200 cases.

If the unit of analysis is the individual study, there are only 24 studies reporting 30-day outcomes for asymptomatics. Using the individual study as the unit of analysis greatly diminishes the effective sample size. Further, because recommended weighting schemes using estimates of variance (cf. Hedges & Olkin, 1985) were not possible, this approach assigns equal weights to all studies.

A major problem with pooling individual cases concerns the possible nonindependence of observations. For example, many of the studies in the database were conducted in a single hospital where facilities and medical staff were the same for each patient. Only 12% of the studies reported CEs across a mix of hospital settings (e.g., university vs. community hospital). Analyses of CE using both Medicare claims data (Fisher et al., 1989) and the present synthesis database

(discussed below) have shown that outcomes can vary systematically by setting.

One solution frequently used (e.g., Hedges & Olkin, 1985) is to employ some type of sensitivity analysis that allows results to be weighted in two or more ways. For example, Table 3 shows outcomes by clinical indication, using the study as the unit of analysis (unweighted) rather than the individual case (weighted). Variations in results did occur, although the differences in point estimates were often small by clinical standards.

If relative rankings are reported by clinical indication for adverse outcomes (Merrick et al., 1986), the hierarchy of best-to-worst outcome proportions shifts when different weighting schemes are used. For example in Table 3, with unweighted outcomes, 30-day (percentage) mortality rates are lower for asymptomatic patients relative to those presenting with a carotid TIA (0.78% and 1.46%, respectively). Unfortunately, when outcomes are weighted, this relationship is reversed (1.28% and 0.90%, respectively). Thus, in this particular instance, the sensitivity approach yields inconsistent results and inferential uncertainty, and the difference may be caused by factors suggestive of bias such as patient selection, surgical skill, hospital setting, and the like. As will be demonstrated below where reporting is adequate, it is possible to estimate the effects of the potential biasing factors.

DISTRIBUTIONAL CHARACTERISTICS OF THE DATA

A second, potential statistical problem concerns the need for a normal distribution of values for individual variables. This criterion is particularly important when multivariate regression analyses are conducted. Using a number of outcome measures as the dependent variable to regress against a number of independent variables calls for the strong assumption of normality for both kinds of variables (Neter, Wasserman, & Kutner, 1985).

As noted by Manning, Duan, and Rogers (1987), however, analyses with epidemiologic data are frequently nonnormal, due in part to the large numbers of zeros. To probe the assumption of normality, a number of univariate analyses were performed for both mortality and morbidity (e.g., incidence of stroke and myocardial infarction). A

series of statistical analyses (Johnson & Wichern, 1982) were then performed that showed those variables were not normal. As a result, square root and inverse sine transformations were performed for the dependent variables when used in multivariate analyses.

ASSESSING THE VARIABILITY OF OUTCOMES

The third general area concerned approaches for explaining variability of outcomes. Conceptually, we were guided by the critical multiplist approach (Shadish, Cook, & Houts, 1986) in which statistical analyses likely to be biased in different directions are used. Here, it also becomes important to determine if existing statistical methods (e.g., regression analyses) are useful for explaining and predicting variance and the policy implications of these methods.

Earlier discussions in this article noted the absence of controlled studies in the CE database and its implications both in terms of the analytic objectives and statistical approaches. Two related analytic strategies were employed to account for different patterns of outcomes from CE surgery.

GENERAL ANALYTIC STRATEGIES

The general analytic strategy involved the use of multivariate regression analysis. The use of multivariate techniques with meta-analytic data has been debated on grounds that the assumption of homogeneity of observation sampling error variances is usually not met in meta-analytic data sets (Hedges & Olkin, 1985). Hunter and Schmidt (1990), on the other hand, note that most statistical tests are robust with respect to violations of this assumption.

Concerns about possible violation of statistical assumptions, coupled with other problems described in this review—missing data, small numbers of study observations, and issues related to true causal relationships—all point to an analytic strategy that does not rely on any single test or approach. Instead, a series of analyses based on different statistical assumptions was employed as part of a general strategy of triangulation or bracketing the actual treatment effect

(Wortman, Reichardt, & St. Pierre, 1978). Each analysis provided some insight into understanding relative risk and assessing the safety of CE surgery. At the same time, no single result could be used as a referent or standard without potential threats to the statistical validity of the finding.

The analytic strategy chosen was one in which multivariate regression models were developed that included one substantive variable reflecting a major variant in the CE surgical procedure. Such variables are common to emerging technologies where the procedure is refined over time (Wortman & Yeaton, 1987). The extent of variation can first be examined through some initial descriptive analyses of main effects. Then, one can begin to investigate relationships and generate additional hypotheses about the relationships among variables. Following that, correlational and multivariate analyses allow variables to be analyzed more completely. Variables can then be added to the regression analysis in a stepwise fashion to examine the independent contribution of individual factors as well as to understand the relative importance of conceptually related groups of variables (e.g., aspects of patient history and risk factors, vessel disease, etc.) beyond those accounted for by confounded methodologic characteristics.

SPECIFYING THE MODEL

This multivariate approach involving an expanded regression model can also be used to assess the relative contribution of important variables. For example, variables measuring the patterns of diagnosis and the surgical procedure itself (e.g., use of surgical shunt in CE) can be added. Finally, once individual variables are examined, clusters of variables can be identified that affect outcomes.

One way to determine if additional variables should be specified (in the regression model) is to examine whether variability in outcomes can be accounted for by sampling error alone (Hunter & Schmidt, 1990). If not, other explanatory variables can then be entered to the analysis. In the case of CE, there was substantial variation in mortality rates for different clinical indications, as shown in Table 3. Consequently, it was not possible to combine outcomes for different surgical indications.

Additional variables were entered into the analysis to test hypotheses related to the generalizability of the findings. These hypotheses were based on relationships found to be true in previously conducted research. For example, patient characteristics such as age have been found to influence CE outcomes (Fisher et al., 1989) and thus were entered into the regression model.

Other researchers have confirmed hypotheses about hospital setting or the evolution of the surgical intervention itself (Wennberg, Freeman, Shelton, & Bubolz, 1989; Wortman & Yeaton, 1987). Table 4 presents data suggesting that variations in patient outcomes exist in different types of hospital settings with 30-day mortality rates substantially lower in university hospitals (i.e., about 25% less). For the subgroup of asymptomatic patients, the 30-day mortality rates are again statistically significantly lower ($p < .01$) in university settings (Langenbrunner et al., 1993). Based on these results, hospital setting was added to the regression model.

Using these results, additional regression and multivariate analyses were performed to explain variations in outcomes for asymptomatic patients (see Table 5). A number of regressions were conducted to assess the amount of variance explained by important predictor variables such as age or gender. This analysis was conducted on both transformed and untransformed dependent variables (T. A. Louis, personal communication, May 14, 1988). These regression analyses used both the 30-day mortality rate aggregated across patient indications as well as the 30-day mortality rate for asymptomatic patients. Two definitions of the shunt variable were included in the analyses (cases A and B in Table 5). In case A, the shunt dummy variable was coded as 1 if the use of a surgical shunt is reported for the majority of the CE patients in an individual study, otherwise it was set at 0. Case B represents the percentage shunt—or the percentage of patients receiving a shunt with CE surgery.

As noted by Sacks et al. (1987) and others, meta-analytic methods often increase statistical power for primary endpoints and subgroups, thereby resolving previous uncertainty or addressing questions not posed in original studies. Although many of the regression coefficients in Table 5 were not statistically significant, two analyses that did emerge as statistically significant were those relating surgical

TABLE 4
Carotid Endarterectomy: 30-Day Patient Status by Hospital Type Across All Clinical Indications

		Patient Status		
Hospital Type	*Total*	*Alive*	*Dead*	*Mortality (%)*
University	28866	28390	476	1.67
Veterans affairs	3637	3563	74	2.08
Community	7764	7605	159	2.09
Mixed	6611	6477	134	2.07
Total	46878	46035	843	1.83

Chi-square = 9.49 ($p < .02$)

TABLE 5
CE for Asymptomatics: OLS Regression Results Using 30-Day Mortality Rate Outcomes[a]

	Asymptomatic Subgroups		*All Clinical Groups*	
Model	*A*[b]	*B*	*A*	*B*
Intercept	.046	.099	.260	−.064
Year	−.002	.002	.037	−.001
Setting	.004	.004	.015	−.016
Gender mix	−.009	.008	−.050	.091
Mean age	−.0005	−.002	.0003	−.002
Shunt	−.0001		−.146*	
Percentage shunt		−.020*		−.0007
F value	0.970	2.63	2.77	0.976
R^2 value	.19	.38	.24	.13

NOTE: Variable definitions are: year = prior to 1985 or 1985 and later years; setting = CE in a university/academic medical center or not; gender mix = percentage of male patients reported in each study; mean age = mean age for patients reported in each study; shunt = surgical shunt reported for majority of CE or not; percentage shunt = percentage of patients receiving surgical shunt with CE.

a. Regression models use 30-day mortality rate as the dependent variable; the unit of analysis is the individual study.

b. Shunt variable included (version A) or the percentage shunt variable included (version B).

*Indicates statistical significance at the $p < .01$ level.

shunt and the proportions of patients receiving shunt to 30-day mortality. In fact, the potential benefit of shunt in improving outcome has been the subject of ongoing debate in the CE literature (cf. Buche et al., 1988).

A strong positive correlation favoring the use of surgical shunt initially emerged in regression models using both 30-day mortality and morbidity outcomes for the subgroup of asymptomatics. The finding also held up in ordinary least squares multiple regression models, using the stepwise function to examine the contribution of individual variables (Langenbrunner et al., 1993). Because the number of usable studies for asymptomatics was also small, the regression model was run on all studies (i.e., irrespective of indication), thereby increasing the number of observations from 32 to 214. The correlation between use of surgical shunt and lowered 30-day mortality increased substantially using this larger set of studies.

DISCUSSION OF CRITICAL ISSUES

Although meta-analytic techniques have typically been used with large numbers of randomized or well-controlled trials, health policymakers too often have few such studies on which to base decisions (Office of Technology Assessment, 1982). This is often true both with new innovations and with those such as CE that have been generally available for many years. To assess the current status of such innovations, it is essential to develop alternative exploratory synthesis techniques that not only help evaluate the validity of the results, but more importantly indicate hypotheses for future study.

This article has reviewed a number of methodological issues that were encountered in performing an exploratory research synthesis on studies with low-quality data. It has addressed a number of these issues concerning missing data, statistical analyses, and modeling. Several points are clear.

First, many issues are amenable to statistical and methodologic bridging despite missing data. For example, missing data on patient sample size was predicted without bias and with high precision from the number of operations performed. Variables can be statistically transformed when values are not normally distributed, allowing further relational analyses.

Second, while some issues cannot be addressed with the same degree of precision, sensitivity analysis reflecting a range of possible values effectively bounds the limits of the estimates. Relatively small

differences in the endpoints of this range increase the confidence in these results. The use of different units of analysis (e.g., weighting by study vs. the individual case) is an example of this approach.

Third, the possibility of establishing relationships when there is a high incidence of missing data awaits further theoretic and methodologic work. For some variables, such as those related to level of occlusion of carotid arteries, the number of missing values was so large that estimation was associated with an unacceptable level of uncertainty.

For the approaches developed and described in this synthesis, a further question remains as to the ideal way to establish their validity. The results of published meta-analyses that include many, relatively small trials have been shown to approximate those of the "gold standard" of large, well-conducted, single studies (Chalmers et al., 1987). Most of the experience with meta-analytic techniques in the biomedical field, however, is with randomized, controlled trials (Sacks et al., 1987); syntheses using quasi-experimental and nonexperimental designs are rare. However, as Longnecker et al. (1988) and Wortman (1994) point out, the application of meta-analytic approaches to nonexperimental data will be valid only if proper precautions are taken to control bias and to ensure that appropriate statistical methods are used.

It is important to note that such exploratory research syntheses alone cannot answer questions regarding efficacy or effectiveness of CE (although standards established by expert panels such as the American Heart Association [1989] can be used as implicit comparisons). Such questions must be addressed by large, multicenter, randomized controlled trials.

Exploratory research synthesis can indicate important hypotheses or variables to include in the design of such large trials. Specifically, the use of a surgical shunt emerged as a significant variable from our research synthesis, but this variable has not been controlled in recent randomized clinical trials such as NASCET. Thus exploratory research synthesis can extend our knowledge about controversial and costly medical procedures and play a crucial role in designing future randomized clinical trials. In this policy-relevant context, exploratory research synthesis methods for handling low-quality data provide an important extension of meta-analysis that enables one to reduce the uncertainty of inference.

REFERENCES

Agency for Health Care and Policy Research (AHCPR). (1991). *Carotid endarterectomy* (Rev. ed.). Rockville, MD: AHCPR. (DHHS Publication No. AHCPR 91-0029)

American Heart Association. (1989). Ad Hoc Committee on Carotid Surgery Standards of the Stroke Council, special report. *Circulation, 79*, 472-479.

Buche, M., Grenier, Y., Schoevaerdts, J. C., Jaumin, P., Ponlot, R., & Chalant, C. H. (1988). Comparative results after endarterectomy of the internal carotid artery performed with or without a shunt. *Journal of Cardiovascular Surgery, 29*, 428-431.

Callow, A. D., Caplan, L. R., Correll, J. W., Fields, W. S., Mohr, J. P., Moore, W. S., Robertson, J. T., & Toole, J. T. (1988). Carotid endarterectomy: What is its current status? *American Journal of Medicine, 85*, 835-838.

Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.

Chalmers, T. C., Berrier, J., Sacks, H. S., Levine, H., Reitman, D., & Nagalingam, R. (1987). Meta-analysis of clinical trials as a scientific discipline: II. Replicate variability and comparison of studies that agree and disagree. *Statistics in Medicine, 6*, 733-744.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Boston: Houghton Mifflin.

Cooper, H., & Hedges, L. V. (Eds.). (1994). *The handbook of research synthesis*. New York: Russell Sage.

Fisher, E. S., Malenka, D. J., Solomon, N. A., Bubolz, T. A., Whaley, F. S., & Wennberg, J. E. (1989). Risk of carotid endarterectomy in the elderly. *American Journal of Public Health, 12*, 1617-1620.

Fuchs, V. R., & Garber, A. M. (1990). The new technology assessment. *New England Journal of Medicine, 323*, 673-677.

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

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

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

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.

ISIS-2 Collaborative Group. (1988). Randomized trial of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected acute myocardial infarction: ISIS-2. *Lancet, ii*, 349-360.

Johnson, R. A., & Wichern, D. W. (1982). *Applied multivariate statistical analysis*. Englewood Cliffs, NJ: Prentice Hall.

Kistler, P. (1989, August). How can we prevent strokes? *Harvard Medical School Health Letter*, pp. 1-5.

Langenbrunner, J. C., Wortman, P. M., Yeaton, W. H., Kshirsagar, A., & Holloway, J. J. (1993). Carotid endarterectomy for asymptomatic patients: Assessing results of a quantitative synthesis. *International Journal of Technology Assessment in Health Care, 9*, 286-303.

Longnecker, M. P., Berlin, J. A., Orza, M. J., & Chalmers, T. C. (1988). A meta-analysis of alcohol consumption in relation to risk of breast cancer. *Journal of the American Medical Association, 260*, 652-656.

Louis, T. A., Fineberg, H. V., & Mosteller, F. (1985). Findings for public health from meta-analysis. *Annual Review of Public Health, 6*, 1-20.

Manning, W. G., Duan, N., & Rogers, W. H. (1987). Monte Carlo simulation evidence on the choice between sample selection and two-part models. *Journal of Econometrics, 35*, 59-82.

Merrick, N. J., Fink, A., Brook, R. H., Park, R. E., Kosecoff, J., Roth, C. P., McCloskey, L., Keesey, J., & Solomon, D. H. (1986). *Indications for selected medical and surgical procedures—A literature review and ratings for appropriateness: Carotid endarterectomy* (Report No. R-3204/6CWF/HF/PMT/RWJ). Santa Monica, CA: RAND.

Merrick, N. J., Fink, A., Park, R. E., Brook, R. H., Kosecoff, J., Chassin, M. R., & Solomon, D. H. (1987). Derivation of clinical indications for carotid endarterectomy by an expert panel. *American Journal of Public Health, 77*, 187-190.

Mosteller, F., Gilbert, J. P., & McPeek, B. (1980). Reporting standards and research strategies for controlled trials. *Controlled Clinical Trials, 1*, 37-58.

Neter, J., Wasserman, W., & Kutner, M. H. (1985). *Applied linear statistical models* (2nd ed.). Homewood, IL: Irwin.

North American Symptomatic Carotid Endarterectomy Trial Collaborators. (1991). Beneficial effect of carotid endarterectomy in symptomatic patients with high-grade stenosis. *New England Journal of Medicine, 325*, 445-453.

Office of Technology Assessment, U.S. Congress. (1982). *Strategies for medical technology assessment*. Washington, DC: U.S. Government Printing Office.

Office of Technology Assessment, U.S. Congress. (1984). *Medical technology and the costs of the Medicare program*. Washington, DC: U.S. Government Printing Office.

Orwin, R. G., & Cordray, D. S. (1985). Effects of deficient reporting on meta-analysis: A conceptual framework and reanalysis. *Psychological Bulletin, 97*, 134-147.

Ozminkowski, R. J., Wortman, P. M., & Roloff, D. W. (1988). Inborn/outborn status and neonatal survival: A meta-analysis of non-randomized studies. *Statistics in Medicine, 7*, 1207-1221.

Roper, W. L., Winkenwerder, W., Hackbarth, G. M., & Krakauer, H. (1988). Effectiveness in health care: An initiative to evaluate and improve medical practice. *New England Journal of Medicine, 319*, 1193-1202.

Sacks, H. S., Berrier, J., Reitman, D., Ancona-Berk, V. A., & Chalmers, T. C. (1987). Meta-analyses of randomized controlled trials. *New England Journal of Medicine, 316*, 450-455.

Shadish, W. R., Cook, T. D., & Houts, A. C. (1986). Quasi-experimentation in a critical multiplist mode. *New Directions in Program Evaluation, 31*, 29-46.

U.S. General Accounting Office. (1992). *Cross design synthesis: A new strategy for medical effectiveness research* (Report No. GAO\PEMD-92-18). Washington, DC: U.S. General Accounting Office.

Venulet, J., Blattner, R., Von Bulow, J., & Berneker, G. G. (1982). How good are articles on adverse drug reactions? *British Medical Journal, 284*, 252-254.

Wennberg, J. E., Freeman, J. L., Shelton, R. M., & Bubolz, T. A. (1989). Hospital use and mortality among Medicare beneficiaries in Boston and New Haven. *New England Journal of Medicine, 321*, 1168-1173.

Wortman, P. M. (1994). Judging research quality. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 97-109). New York: Russell Sage.

Wortman, P. M., Reichardt, C. S., & St. Pierre, R. G. (1978). The first year of the Education Voucher Demonstration: A secondary analysis of student achievement test scores. *Evaluation Quarterly, 2*, 193-214.

Wortman, P. M., & Yeaton, W. H. (1987). Using research synthesis in medical technology assessment. *International Journal of Technology Assessment in Health Care, 3*, 509-522.

Yeaton, W. H., & Wortman, P. M. (1993). On the reliability of meta-analytic reviews: The role of intercoder agreement. *Evaluation Review, 17*, 292-309.