The Clinical Neuropsychologist

ISSN: 1385-4046 (Print) 1744-4144 (Online) Journal homepage: http://www.tandfonline.com/loi/ntcn20

Meta-Analysis in Neuropsychology: Basic Approaches, Findings, and Applications

George J. Demakis

The Clinical Neuropsychologist, 20: 10–26, 2006

ISSN: 1385-4046
DOI: 10.1080/13854040500203282

Psychology Press
Taylor & Francis Group

META-ANALYSIS IN NEUROPSYCHOLOGY: BASIC APPROACHES, FINDINGS, AND APPLICATIONS

George J. Demakis
Department of Psychology, University of North Carolina at Charlotte, Charlotte, NC, USA

This article is a broad review of the approaches, findings, and applications of meta-analyses on clinical neuropsychological topics. The review is divided into four sections: basic characteristics of meta-analysis; the value of meta-analysis for neuropsychological investigations; illustrative findings from various meta-analyses on neuropsychological topics demonstrating the type of questions that can be answered; and problems and limitations of meta-analysis with a focus on future directions. The article is not intended to exhaustively review all the relevant literature nor detail the technical aspects of meta-analytic techniques, but rather it is designed to provide a basic, conceptual introduction for the general reader, particularly the clinician, to aid in comprehension of the now burgeoning literature that uses meta-analysis. Throughout, illustrative examples are developed and references are made to the practice of clinical neuropsychology.

INTRODUCTION

Given the different theoretical models, wide diversity of research designs, and conflicting findings that characterize many literatures within psychology, it is not surprising that methods of integrating or synthesizing research have long been considered important. The traditional narrative review, in which broad conclusions are developed based on a careful reading of the extant literature, has been the workhorse of these types of studies. However, significant critiques have been leveled against such approaches, including, for instance, an overreliance on published studies and an insensitivity to power and sample size issues. These issues, particularly the latter two, have been addressed by meta-analysis, a relatively new method of research synthesis now frequently employed in psychology. The first modern meta-analysis was Smith and Glass's (1977) groundbreaking work on the efficacy of psychotherapy. Their findings were based on aggregation of 375 psychotherapy outcome studies and challenged Eysenck's position that psychotherapy was ineffective. Eysenck (1978), in an early critique of meta-analysis, was unconvinced of its value and considered the method "an exercise in mega-silliness" (p. 517). Since then, despite continuing concerns about the validity of and necessity of meta-analytic approaches (Feinstein, 1995; Sharpe, 1997), they have been widely used in many areas of

Address correspondence to: George J. Demakis, University of North Carolina at Charlotte, Department of Psychology, 9201 University City Blvd., Charlotte, NC 28223-0001, USA. Tel.: 704-547-4750. Fax: 704-687-3096. E-mail: gdemakis@ email.uncc.edu

Accepted for publication: November 11, 2004.

psychology, including neuropsycholology, and are now generally considered an acceptable method of data synthesis.

The purpose of this article is fourfold. First, some of the basic characteristics of meta-analytic approaches are highlighted. Second, reasons for the particular relevance and suitability of meta-analysis for neuropsychological research are discussed. Third, illustrative findings from studies that have used these methods in neuropsychology are reviewed and, more broadly, the sorts of questions that can be answered in the field with this approach are addressed. Finally, limitations and problems of using these techniques are discussed, as well as some future directions that may be profitably pursued. Within each of these sections my goal is to address issues relevant for the clinician both in terms of actual practice and how to read the relevant literature in a conceptual, non-technical fashion. This review is not meant to (a) exhaustively review all the meta-analyses that have been conducted in neuropsychology, (b) address every technical or statistical issue, or (c) to provide the "hands-on" information necessary to actually perform a meta-analysis. Rather, it is designed to be a user-friendly introduction for the clinician regarding how meta-analysis can inform practice and, ideally, provide assistance to comprehend the increasing number of meta-analyses related to clinical neuropsychology. For the reader interested in more technical issues, several excellent books are highly recommended (Cooper & Hedges, 1994; Hedges & Olkin, 1985; Hunter & Schmidt, 1990; Lipsey & Wilson, 2001). Other useful resources include articles on how to write meta-analytic reviews (e.g., Rosenthal, 1995) and a guide for judging the scientific quality of meta-analyses (Jadad & McQuay, 1996).

WHAT ARE THE DEFINING CHARACTERISTICS OF META-ANALYSIS?

Though there are several different meta-analytic methods and statistical approaches (see Bangert-Drowns, 1986), they share several general characteristics. First, meta-analysts typically seek out all relevant studies on a topic, whether published or unpublished, to avoid what has been termed the publication bias. This bias has been documented in psychological and medical literatures (for review see Begg, 1994) and occurs when research with statistically significant findings is more likely to be published than that with null findings. As a result, reviews that include only published results are likely to present a biased and misleading picture of the research. Second, inclusion/exclusion criteria, driven by research questions, are formulated to determine which studies are ultimately analyzed. These explicit criteria provide discipline to this early stage of the research process and a transparent account of how the studies are selected. Ideally, the selected studies are conceptually comparable, deal with similar issues and constructs, and provide enough statistical information to extract effect sizes. Third, the selected studies are quantitatively analyzed so that an effect size, typically the difference between the mean of two groups divided by a standard deviation measure in neuropsychological research, is obtained. Effect sizes are thus expressed in the common metric of standard deviation units which, because of this standardization, can be averaged across studies. Consequently, a precise quantitative analysis of the difference between groups (i.e., the effect size) across many studies comprised of many individuals can be computed. Some meta-analysts also (a) weight the effect sizes by sample size, so that studies with larger numbers of participants have relatively greater influence on the final effect size, (b) assess the variability or homogeneity of the final mean effect size to determine if the contributing effect sizes estimate a common population mean, and (c) correct effects sizes to account for various forms of bias or unreliability (see Hunter & Schmidt, 1990). Finally, the relationship of the effect size to various characteristics of the studies, such as patient demographics or study quality, can be determined. These so-called moderator variables are typically coded when the selected studies are analyzed and can provide a rich empirical description of the relationship between the effect size and other variables of interest. Moderator analyses are more difficult to perform when the number of analyzed studies is small or when the candidate variables are insufficiently described.

The above characteristics suggest some of the potential advantages of meta-analysis compared to the traditional narrative review in psychology. I address only two specific advantages here, as they are particularly relevant for neuropsychological research. Perhaps the most appealing aspect of meta-analysis is that it is not dependent on traditional null hypothesis significance testing (see Kazdin, 1998, pp. 373–376, for full description of the difficulties with this approach). In this scenario, one rejects the null hypothesis when (typically) $p < .05$, with p dependent on both the strength of the relationship between variables (i.e., effect size) and sample size. Statistical tests conducted on groups with small sample sizes are thus likely to be underpowered and are less likely to reach statistical significance (see Cohen, 1992). Because traditional narrative reviews may evaluate a body of literature simply by comparing the number of studies that obtain statistical significance vs those that do not (i.e., vote counting), they can provide misleading conclusions, particularly when many of the studies are underpowered. Moreover, the yes–no binary decision-making that characterizes this approach provides minimal information about the relationship between variables. Meta-analysis does not suffer from these problems. Rather than considering statistical significance, meta-analysis analyzes effect sizes and provides a precise measure of the strength of association between variables or the difference between groups. Conventionally, .2 or less has been considered a small effect size, .5 a medium effect size, and .8 or larger a large effect size. A second advantage of meta-analysis is the ability to detect relationships between variables that tend to be obscured within the literature. Because many individual study characteristics or moderators are coded and then analyzed for each study, sophisticated relationships among variables can potentially be examined. Even in well-conducted narrative reviews, it is extraordinarily difficult to consider all relevant variables and their potential relationships—this is especially so when dealing with large, complex literatures.

WHY USE META-ANALYTIC TECHNIQUES IN NEUROPSYCHOLOGY?

First, like many areas in psychology, neuropsychology is marked by diverse and, at times, contradictory findings; individual studies may vary on a bewildering array of variables including the "usual suspects" of patient age, gender, education, premorbid functioning, and emotional status that have been of long-term interest in neuropsychology. Differences in injury characteristics, etiology, tests administered, and even the methods of scoring tests are other, but surely not the only, complicating variables. Anyone who has reviewed a body of neuropsychological literature can attest to the ''messiness'' of the collective findings—as noted above, one of the advantages of meta-analysis is that it can meaningfully organize the literature and evaluate the impact of moderators on the variable of interest.

Second, certain neuropsychological studies and research areas tend to be underpowered, particularly when relatively rare disorders are studied or when exclusionary criteria are stringent. In traditional hypothesis testing, this means that statistical tests do not correctly reject false null hypotheses due to insufficient sample size or, more simply, the tests are unable to detect an effect that is in fact present. For example, in a meta-analysis on the sensitivity of the Wisconsin Card Sorting Test (WCST) to frontal lobe damage, I included studies that compared participants with exclusively frontal damage to those with exclusively non-frontal damage (Demakis, 2003). If a conservative small-to-medium effect size ($d = .3$) is assumed between these groups on this test, only 3 of the 26 comparisons had sufficient power to reject correctly the null hypothesis at $p < .05$. The remaining comparisons were underpowered and unable to detect such a small effect. While the issue of statistical power is of broad concern in psychology and several studies have demonstrated that much research tends to be underpowered (Bezeau & Graves, 2001; Rossi, 1990), the increasingly refined questions neuropsychologists are asking make this issue more acute. For example, within the frontal lobe damage/executive processing literature, researchers are increasingly studying regional variations within the frontal lobes (e.g., dorsolateral vs. orbitofrotnal) rather than aggregating these various subregions into a ''lobe.'' Donald Stuss and colleagues (e.g., Stuss & Levine, 2002) are consistently making such precise neuroanatomical determinations within frontal regions and, in doing so, have helped to elucidate regional functional differences within these areas. Yet, because of the difficulty in obtaining a large number of participants with, for example, selective left dorsolateral damage, these studies tend to be characterized by small numbers of participants. The resulting low power makes reliance on traditional statistical significance testing potentially misleading. Meta-analysis is useful here, as noted above, because it is not influenced by sample size—an effect size difference between groups is generally the same in groups of 10 or 100 participants and conveys the same relationship between variables.

Third, as meta-analysis quantitatively determines differences between groups in the form of an effect size, much corresponding useful information can be obtained to aid the clinician. Assuming similar base rates, one method has been to convert the effect size to a Z score and then percentile. For instance, if the difference between Group A and Group B is $d = .7$ in favor of Group A, it indicates that the mean of that group is 7/10 of a standard deviation higher. From a Z table it can be determined that this effect sizes translates into the 76th percentile, indicating that the average subject in Group A is better than 76% of the individuals in Group B. In a more recent approach, effect sizes are converted to overlap percentages in which the distributions of the two groups can be graphically plotted and compared. For instance, an effect size difference of 1.0 (i.e., one standard deviation) between two groups corresponds to an overlap percentage of 44.6%, meaning that almost half of the distribution of the scores of both groups overlap and only 55.4% (100% – 44.6%) do not overlap. A larger effect size is associated with a smaller overlapping percentage and hence an enhanced ability to discriminate between groups (Zakzanis, 2001).

Zakzanis has suggested that an effect size of $d = 3.0$, with a corresponding overlap of 5%, would be appropriate for use as a clinical marker of neuropsychological disorders or diseases. In comparison, a traditional narrative review of the same issue might only conclude that the groups do in fact differ statistically, but typically would not address the magnitude of this difference or how well groups could be discriminated. The value of meta-analysis to quantify differences between groups is obvious here, as this can be translated to clinically meaningful information.

A final advantage, one more of orientation rather than of method, is the manner in which the meta-analyst approaches and analyzes studies. Meta-analysis requires reading and evaluating studies with an emphasis on the data, not on what is reported or argued about these findings. Rosnow and Rosenthal (1989) consider this a "new intimacy" between the reviewer and the published literature. Such intimacy requires the careful reading of many articles to determine if they meet the inclusion criteria and, if so, the subsequent computation of an effect size. Rosnow and Rosenthal playfully term the hard-driving detective work required of meta-analysis the "decrease in the splendid detachment of the full professor." No longer can research assistants obtain and/or summarize the literature as is done in traditional narrative reviews. This orientation is similar to the way in which clinical neuropsychologists examine (or should examine) individual patient test data. What is key is not necessarily the written report, but rather the data, including actual patient responses and test scores, and how well such data support the conclusions. This is particularly true when evaluating or using the work of other practitioners, but also in comprehending one's own work. For instance, when evaluating a patient multiple times, neuropsychologists tend to compare the obtained scores (i.e., how many items were recalled on a particular test at each evaluation), rather than only considering the qualitative description (e.g., performance within the average range). It is not a stretch to suggest that neuropsychologists, more so than other clinical psychologists, are more concerned in general with norms and norm development, test scoring and administration accuracy, and score interpretation. Such "data intimacy" parallels the orientation of the meta-analyst.

HOW TO DO A META-ANALYSIS IN NEUROPSYCHOLOGY: CONCEPTUAL NUTS AND BOLTS

By this point, I hope to have convinced the reader that meta-analytic techniques are natural to neuropsychology and similar, in many ways, to the ways in which clinicians interpret and understand data. Within this section, selective issues are addressed that can guide individuals interested in performing a meta-analysis on a neuropsychological issue or topic. This is not meant to be a technical or statistical review, as this is beyond the scope of this article, and others have amply addressed these issues, but rather to highlight unique issues and problems likely to face the neuropsychological researcher when conducting such a study. As an additional preface, judgment calls are necessary at each stage of this research process (Wanous, Sullivan, & Malinak, 1989); these will, in part, determine the quality and impact of the particular meta-analysis. To meaningfully make such decisions, the researcher must comprehend meta-analytic techniques and procedures, but also have a broad, yet intimate familiarity with the relevant research. Ideally, these judgments should be reported within the research report and, where feasible, the effects of these decisions empirically evaluated, perhaps as moderator variables.

Problem Formulation

Meta-analyses are typically performed in research areas that have diverse, contradictory, and/or mixed findings. Neuropsychology is no stranger to such bodies of literature. Within neuropsychology, the most compelling meta-analyses have tackled literatures with such diversity, as well as controversial problems or theoretical differences between researchers. For instance, Binder, Rohling, and Larrabee's (1997) meta-analysis on the controversial neuropsychological effects of mild TBI found only relatively minor deficits when such individuals were compared to controls. Demakis (2003) evaluated the sensitivity of the WCST to frontal lobe damage and found, contrary to narrative review articles (e.g., Mountain & Snow, 1993) that this measure is sensitive to such damage, though sensitivity varies by moderators. Heinrichs and Zakzanis (1998) found a broad pattern of cognitive deficits in schizophrenia, rather than specific executive deficits that might support a model of frontal lobe dysfunction, proposed by some, in this disorder. These types of meta-analyses are likely to be well received as they can help resolve and answer current issues and questions, advance model/theory building, inform clinical practice, and provide directions for future research.

Study Retrieval

Once one has selected a research topic or problem, the next step is to determine the studies to include or exclude and, more broadly, where to find such studies. Databases rich in neuropsychological research include the obvious PsychInfo and Medline, and to a lesser extent Dissertations Abstracts International. Reading reference lists of relevant articles, hand searching through relevant journals, and talking to colleagues to obtain unpublished data are necessary methods for obtaining studies. Given that many clinical neuropsychologists keep their own databases, this can also be a source of unpublished data. Because of the potential of publication bias, it is important to seek out unpublished data, whether in dissertation form or from other researchers or colleagues.

Inclusion/Exclusion Criteria

Inclusion/exclusion criteria are an integral part of the meta-analytic process and are determined by the questions the researcher asks, as well as the current state of the literature. To accurately develop such criteria, the meta-analyst must have a thorough familiarity with the literature, including both conceptual and practical issues, so that only the studies that appropriately address the research question are included. Of course, these judgments can be difficult and the source of reasonable disagreement among researchers. In my own research on the WCST (Demakis, 2003), I chose not to include studies that used participants with anterior communicating artery aneurysm in the frontally damaged group, given the controversy about whether these individuals have suffered frontal or more diffuse brain injury. My goal, in fact, was to select a group of studies that assessed patients with frontal and only frontal damage—I excluded the aneurysm participants because they potentially compromised the integrity of the frontal group. Though it is possible that another meta-analyst would have made a different decision regarding these participants, the decision was nonetheless clearly argued and reported in the manuscript. Thus, as my own study illustrates, meta-analysis provides a transparency to the research process not traditionally observed in narrative reviews.

One particular inclusion/exclusion criterion that deserves special mention is determination of the proper control or comparison group for use in meta-analysis. In a meta-analysis on Alzheimer's disease, should one include studies that compare these individuals to a non-neurologically impaired control group or perhaps to studies that compare them to a different neurologically impaired group or a psychiatrically involved group? These latter studies are likely to be more clinically relevant, as one is rarely asked to determine if an individual has Alzheimer's disease or is entirely normal, but the effect sizes are likely to be smaller as the clinical groups are likely to have cognitive deficits. For instance, Johnson-Selfridge and Zalewski (2001) found an effect size of $d = -1.45$ when studies compared individuals with schizophrenia to normal controls, but a smaller effect size of $d = -.40$ when these same individuals were compared to other psychiatrically involved individuals. Both effect sizes indicated worse performance by the individuals with schizophrenia. While there is no hard and fast rule for this important judgment call, the use of the comparison group should be guided by practical considerations (e.g., what types of studies exist in the literature?) and one's research question. If one is interested in knowing whether neuropsychological impairment exists in a population and/or the magnitude of such impairment, a normal comparison group may be more appropriate. In cases driven by clinical considerations, such as how well do certain tests discriminate between dementia and depression in the elderly, use of such clinical groups seems more appropriate.

Selection/Combination of Dependent Variables

A major challenge in meta-analysis is that studies tend to compare groups on multiple variables often assessing similar constructs (e.g., verbal memory as assessed by the California Verbal Learning Test and the Words subtest of the Recognition Memory Test). Should these variables be combined or analyzed separately? The complexity is amplified when other studies use dissimilar variables, such as the Wechsler Memory Scale Logical Memory and Verbal Paired Associates subtests, to tap verbal memory. There are several ways of managing this issue (Heinrichs & Zakzanis, 1998). These variables can be (a) combined rationally based on theoretical assumptions about what they measure, (b) combined empirically based on factor-analytic findings, or (c) they can be analyzed separately. Currently, the vast majority of meta-analyses in neuropsychology combine the variables rationally in a face valid method into well-known constructs, such as attention and verbal memory. Meta-analysts here are simply reflecting what is typically done clinically and how reports tend to be conceptualized. However, it is preferable that these decisions be made on an empirical basis, perhaps with the aid of factor analysis, such that only similar variables are aggregated together. Such an approach goes toward obviating what has been termed the "apples and oranges" problem in meta-analysis (see below).

While important factor analytic work has been published (e.g., Boone, Ponton, Gorsuch, Gonzalez, & Miller, 1998; Leonberger, Nicks, Larrabee, & Goldfader, 1992), this has not been done widely with various neuropsychological measures nor has this been done widely in disparate populations. It is unclear, for instance, whether findings from Boone et al.'s factor analysis of executive measures would "hold up" in populations other than the mixed neuropsychiatric group they studied. Despite this potential problem, this is a more defensible approach then simply collapsing tests based on rational groups. As a final note, some meta-analyses in neuropsychology have not aggregated different tests of the same construct, but have evaluated each separately. Though this has merit, as it provides a precise analysis of a specific test, such analyses tend to suffer from a small number of contributing effects sizes for many tests and they can become unwieldy and difficult to interpret.

Basic Analytic Techniques

There are two broad classes of statistical techniques used to compute effects sizes in meta-analysis: correlational analyses (expressed as r) and standardized mean effects, typically the difference between group means divided by some expression of the standard deviation (often expressed as d). Though they are similar, as they both aggregate multiple studies and can easily be converted to one another, the d family of analyses is most widely used in neuropsychology (but see Wishart & Sharpe, 1997, for an exception). Such usage reflects the method of reporting data in most studies and parallels the customary thought process of clinicians who tend to think about test performance in terms of standard deviations (i.e., how far below normative expectations is this performance?). Because it is also easier to determine overlap percentages from d vs. r values, use of this measure of an effect size is recommended for neuropsychological topics. In addition to the issue of which measure of effect size to use, there are several other important analytical issues that should be considered in meta-analysis but are beyond the scope of this article. These include the following: modeling effect sizes with either fixed or random effect models (Hedges, 1994; Raudenbush, 1994), computation of effect sizes from correlated or repeated measure designs (Morris & DeShon, 2002), and the use of Bayesian methods in meta-analysis (Louis & Zelterman, 1994).

CONTRIBUTIONS OF NEUROPSYCHOLOGICAL META-ANALYSES

In this section some of the important contributions meta-analyses have made in neuropsychology are reviewed, particularly as related to clinical issues and problems. This review is not meant to be exhaustive, but rather it is designed to highlight key, illustrative findings. The discussion is framed by four core questions.

Does Neuropsychological Impairment Exist? If So, What Is Its Magnitude?

Meta-analytic techniques provide a powerful method for addressing these questions as they can precisely quantify differences between groups. An excellent example is the Binder et al. (1997) study that examined the neuropsychological sequalae of mild TBI. To address this controversial issue, they included only studies that evaluated individuals with a history of mild TBI who had not presented for treatment of that injury. They found minimal differences between these individuals and matched controls without injury on many neuropsychological measures. The only significant difference was on a composite measure of attention ($d = -.2$), with worse performance for mTBI individuals); the difference was so small that it was judged to be virtually indistinguishable clinically (i.e., it was not clinically meaningful). Similarly, Grant, Gonzalez, Carey, Natarajan, and Wolfson (2003) found that chronic marijuana users did not differ from matched non-user controls on most neuropsychological domains. The only differences, which were again relatively small (approximately $d = -.20$), were that marijuana users performed more poorly on the learning and forgetting composite measures. Reger, Welsh, Razani, Martin, and Boone (2002) meta-analytically examined the neuropsychological functioning of participants across four groups of the human immunodeficiency virus (HIV) spectrum (HIV–, HIV+ asymptomatic, HIV+ symptomatic, and AIDS). Interestingly, a dose–response–like effect was found across a broad range of neuropsychological domains; relatively small effects sizes were found for the HIV– vs. HIV+ symptomatic comparison, but large effects for the HIV– vs. AIDS comparison. In a more traditional assessment of dose–response effects, Meyer-Baron, Schaeper, and Seeber (2002) found a correlation ($r = .26$) between the amount of mercury exposure in occupational workers and the effect size difference between neurobehavioral functioning in mercury exposed and non-exposed groups.

What Is the Nature or Signature of Neuropsychological Impairment?

Once differences between groups are quantified, meta-analysis allows for description of the pattern of differences between groups that can inform theory and/or clinical practice. For instance, in the Reger et al. (2002) study mentioned above, evidence suggestive of a subcortical dementing process in HIV/AIDS (e.g., such as larger effect sizes for mental/psychomotor speed and executive functioning vs other neuropsychological domains) was found. In another comprehensive meta-analysis, Heinrichs and Zakzanis (1998) examined the neuropsychological impairment in individuals with schizophrenia. As compared to non-psychiatrically ill individuals, they performed more poorly across a wide-range of constructs; the largest difference was for global verbal memory ($d = -1.41$) and the smallest was for Block Design ($d = -.46$). Yet, there was no evidence for a selective impairment in executive processing measures, as might be predicted with the presumed frontal lobe dysfunction in schizophrenia. In a different type of meta-analysis, Schretlen and Shapiro (2003) found that, not surprisingly, cognitive functioning was significantly more impaired in moderate-severe traumatic brain injury compared to mild traumatic brain injury ($d = -.74$ and $d = -.24$, respectively). Interestingly, performance essentially returned to baseline in mild injury within 1–3 months, but the moderate-severe group still evidenced marked cognitive impairment over 2 years post-injury.

Are Neuropsychological Tests Valid?

Through quantification and the ability to specifically compare groups across various measures, meta-analysis affords a unique method of evaluating test validity.

For instance, I conducted a meta-analysis on the validity of various executive processing measures by comparing performance of frontally damaged individuals to nonfrontally damaged individuals (Demakis, 2004). The findings were mixed; frontally impaired participants performed worse on Trails A and all components on the Stroop test, but there were no differences Trails B or the Category test, putative measures of frontal lobe functioning. Effect size differences between groups were also relatively similar for all components of the Stroop test, which indicates that the frontally damaged participants did not perform selectively poorly on the Color-Word subtest, as might be predicted from traditional conceptualizations of the test. In all, these results suggest that some presumed frontal lobe tests are not valid measures of frontal lobe functioning and that other tests may be sensitive to frontal lobe impairment in ways other than traditionally conceived.

In a broad evaluation of test validity, Vickery, Berry, Inman, Harris, and Orey (2001) evaluated the sensitivity of various malingering tests to malingering, whether in analogue designs or actual clinically suspected or identified malingerers. This meta-analysis found a relatively large overall effect size ($d = 1.13$) for the difference between malingerers and individuals exerting sufficient effort, but considerable variability across tests. The Digit Memory Test, 21-Item Test, and the Portland Digit Recognition Test demonstrated larger effect size differences than the 15-Item Test and the Dot Counting Test. When the types of studies were assessed, analogue designs yielded higher effect sizes than did designs that used suspected clinical malingerers. In a similarly broad study, Christensen, Hadzi-Pavlovic, and Jacomb (1991) found that memory tests, as compared to language, praxis, or perception tests, best differentiated individuals with dementia from healthy elderly participants. When dementia screening instruments were evaluated, the Mini-Mental State Examination (MMSE) was more sensitive across all levels of dementia than other measures (e.g., the Blessed Dementia Scale). In other words, effect size differences between dementia and healthy elderly participants were larger for the MMSE than the Blessed Dementia Scale ($d = -2.91$ vs. $d = -2.48$). When converted to overlap percentages, this one-half standard deviation difference reflects a larger spread between the scores and thus enhanced separation of groups with the MMSE. In all, the above studies illustrate that validity issues can be addressed by meta-analytically evaluating test performance in clinical populations and by comparing each test's discriminability in ''head-to-head'' comparisons.

There are additional ways to evaluate validity issues such as using meta-analysis to compare the accuracy of neuropsychological testing to other methods or techniques. Zakzanis (1998), for instance, compared studies that used either neuropsychological testing or neuroimaging to compare individuals with Alzheimer's disease to normal controls. Interestingly, the effect size difference for neuropsychological testing (i.e., the California Verbal Learning Test and the Wechsler Memory Scale—Revised combined) was much larger ($d = 3.2$) than magnetic resonance imaging (MRI: $d = 1.4$) or positron emission topography (PET: $d = 1.2$). All of these effect sizes favored the normal controls. In a study that only compared methods of neuroimaging frontal brain regions, Zakzanis and Heinrichs (1999) compared structural vs functional methods in discriminating between individuals with schizophrenia and normal controls. When brain volume was assessed, either via computed topography (CT) or MRI of the frontal brain, the effect size was .36, but when PET technology was used, effect sizes were .64 at rest and 1.13 when activated. Again, all of these effect sizes favored the normal controls.

Are Moderators Important for Understanding Effect Sizes?

In addition to differences between groups, meta-analysis can be used to investigate the relationship between study moderators, such as patient characteristics or study design, and the resulting effect size. For instance, I found time since injury and method of WCST administration important for understanding the relationship between frontal lobe damage and WCST performance (Demakis, 2003). A significantly larger effect size was obtained when participants were within one year of their injury vs one-year post-injury and when Nelson's WCST modified administration method was used rather than Heaton's standard method. Kindermann and Brown (1997) also found moderators important in a meta-analysis that compared memory performance in the depressed elderly with age-matched non-depressed individuals. The overall mean effect size difference was −.60 with poorer performance by depressed individuals; examination of moderators, however, found that this effect size varied based on a patient characteristic (whether the studies' patients had unipolar or mixed depression) and a study characteristic (whether depressed patients were recruited from a patient list for treatment or whether they were called or otherwise contacted). Effect sizes were significantly larger when studies used mixed depressed individuals and when patients were recruited from a patient list for treatment. Finally, Johnson-Selfridge and Zalewski (2001) found that effect size differences on executive functioning measures between individuals with schizophrenia and non-psychiatrically impaired controls was related to several key illness characteristics. Positive and negative symptoms, as well as the number of hospitalizations, were correlated with effect size, but there was no relationship for duration of illness, Brief Psychiatric Rating Scale, and chlorpromazine equivalence. The above studies attest to the importance of analyzing moderator variables, as they can be critical to comprehending the complicated ways in which variables may influence effect sizes.

PROBLEMS AND LIMITATIONS

Despite the important role for meta-analysis in synthesizing and integrating neuropsychological research, it would be naïve to assume that these approaches are without their own unique set of issues and problems. In fact, since Eysenck's (1978) early critique, there have been broad attacks against both the basic analytical and conceptual framework of meta-analysis (see, for example Sharpe, 1997). Several reviews have fairly and accurately discussed threats to the validity of meta-analysis (Cooper & Hedges, 1994; Lipsey & Wilson, 2001, pp. 7–10). Review of all of these issues is beyond the scope of this article, but below three key issues and their relevance for meta-analytic studies in neuropsychology are discussed.

The "Crowding Out Wisdom" Problem

Some have argued that the structured and somewhat mechanical approach to coding and analyzing studies can result in a synthesis that is not fully sensitive to important issues, such as the social context of the research or theoretical influences and implications (Feinstein, 1995). Like the neuropsychological report that "loses" the patient in a sea of cognitive data rigidly reported and interpreted, the meta-analyst may review the literature in such a technical fashion that broad themes or "big picture" issues are missed. Traditional narrative reviews are presumably less likely to suffer from this problem. Despite this concern, many meta-analyses tend to include both meta-analytic findings, as well as components of the traditional qualitative reviews—these approaches need not be mutually exclusive. While there is obviously variability in meta-analysts skill in doing so, there have been many excellent meta-analyses within neuropsychology that have interpreted their findings within more narrative-like reviews that addressed broad theoretical and neuroanatomical issues. This was the case, for instance, in meta-analyses by Zakzanis and Heinrichs (1999) on schizophrenia and executive deficits and Meiran and Jelicic's (1995) on implicit memory in Alzheimer's disease. Such well-executed meta-analyses do not lose sight of their place within the broader literature and, in fact, their success depends, at least partially, on their ability to read that literature in a meaningful fashion and to understand the larger context of their work.

The "Apples and Oranges" Problem

One of the most enduring criticisms of meta-analysis is that multiple different measures are combined into one group. This "apples and oranges" problem would be reflected in the combination of various memory tests (e.g., the California Verbal Learning Test, Recognition Memory Test, and Logical Memory and Word Pairs from the Wechsler Memory Scales) into one broader category of "verbal memory." Though similar, does it make sense to include these measures together? How should performance on this broad and heterogeneous category of memory functioning be interpreted? While this critique does have some validity, a reasonable method of addressing this issue is to use empirical methods to drive how variables are combined. As mentioned earlier, this can be achieved via factor analysis of neuropsychological tests; ideally, variables that load on the same factor can be combined into a single group. Rohling and Demakis (2006) used this approach, when possible, in their meta-analysis on mercury exposure and included all components of the Stroop test in the processing speed category based on the factor analytic work of Boone et al. (1998). This was done despite the fact that the Color-Word component of the test has typically been considered a measure of executive functioning. While such an approach is currently limited, as factor analyses may not have been done in the same populations as they are applied to or with the same tests, it is nonetheless likely to create "purer" categories of tests vs rationally grouping tests based on long-held assumptions about what the tests purportedly measure. Later interpretation and application of the resulting findings is also likely to be more accurate. The interested reader is referred to factor-analytic research to guide how variables should be combined (e.g., Boone et al., 1998; Larrabee & Curtiss, 1995; Leonberger et al., 1992).

If "pure" groups of variables cannot be created, another approach to dealing variability or heterogeneity in effects sizes is via moderator analyses (see Lipsey & Wilson, 2001). After initial analyses marked with considerable heterogeneity, similar variables can be "pulled out" or analyzed separately. For instance, after analyses of a broad verbal memory factor, individual effect sizes comprising aspects of verbal memory (e.g., recall and recognition) can be separately analyzed. In this way, the variables are grouped into purer measures of the construct of interest.

THE "GARBAGE-IN, GARBAGE-OUT" PROBLEM

Because meta-analysis typically seeks to comprehensively review all the relevant literature, studies of varying methodological quality are likely to be included. Within the psychotherapy outcome literature, methodological quality typically surrounds subject selection, with higher quality studies using random assignment to groups and lower quality studies using non-random selection methods. Echoing the apples and oranges problem, does it make sense to include these studies together into one analysis? Will the various methodological flaws of the poorer quality studies result in a misleading and inaccurate portrayal of the higher quality studies and, more broadly, the literature? While these questions raise a variety of fairly complicated conceptual and statistical issues, the associated quality issues are usually handled in one of two ways. Studies of lower quality can be excluded via development of strict inclusion criteria, despite the loss of potentially valuable information and the ability to generalize findings broadly, or all studies can be included and study quality can then be evaluated as a potential moderator variable. In the latter scenario, effect sizes can be empirically compared for studies judged to have low vs high quality; this approach has the advantage of retaining all the relevant studies and potentially elucidating how methodological variation contributes to study differences. Despite this appeal, making judgments about a study's quality can be troublesome and can introduce another source of unreliability to the meta-analysis.

Because there are relatively few treatment studies in neuropsychology (but see meta-analysis on the efficacy of cognitive rehabilitation by Park and Ingles, 2001), quality issues rarely surround random vs non-random assignment as they traditionally have in psychotherapy or medical treatment literatures. Rather, other aspects of quality have been evaluated such as how well participants in different groups were matched on education (Kindermann & Brown, 1997) and the improved spatial resolution over time of neuroimaging techniques in depicting the neuroanatomical correlates of Alzheimer's disease (Zakzanis, Graham, & Campbell, 2003). Kindermann and Brown, for example, found that the effect size difference for memory between elderly controls and those with depression was significantly higher in studies that poorly matched groups by education vs studies that matched groups reasonably well on this variable ($d = -.78$ vs. $d = -.45$, respectively). The poorly matched studies suggest that education may have confounded the relationship between depression and memory functioning. As this study illustrates, the use of quality-related variables as moderators can assist in more clearly comprehending methodological issues within a literature and ideally provide an empirical guide for future research.

FUTURE DIRECTIONS

Based on the increasing number of meta-analyses being done in psychology and neuropsychology every year, it is clear that these methods have been broadly accepted and are likely to remain as important methods of research synthesis. There are, however, several approaches and topics within neuropsychology that remain relatively underinvestigated with meta-analysis and could benefit from future inquiry. Three of these are described below to illustrate where the field may profitably move in the future.

Test Validity

Meta-analysis is a relatively underutilized approach to establishing and/or comparing test validity. For instance, if one is interested in comparing the validity of various verbal memory tests, one can compare how well they discriminate between groups known to have such deficits (e.g., left temporally damaged individuals) vs controls. Effect size differences can be computed across tests—the comparison with the largest effect size difference (assuming all other aspects of the studies are similar) is the most sensitive to such damage and to verbal memory impairment. This approach, which can be considered a known-groups design (i.e., one of the groups definitely has the disorder or damage that one is investigating), has not been widely used to address validity issues.

Comparing across Symptom Domains

Current meta-analyses in neuropsychology typically focus on only one symptom domain, usually cognitive or psychological functioning. Few have evaluated multiple domains within the same study (but for exception, see Rohling and Demakis, 2005). This is unfortunate because the quantification of meta-analysis is uniquely suited to make such comparisons across broad symptoms domains. Differences between such domains can ultimately provide important clinical and theoretical information about a disorder. For instance, how do effect sizes differ for motor and cognitive symptoms in Parkinson's disease? Do these effect sizes change across the disease process and, if so, how? Are moderators (e.g., medications, gender, education) differentially related to these domains? Another study of broad interest, particularly for those in forensic circles, would be comparison across neuropsychological and psychological domains in individuals with mild TBI. If there are differences across these domains, do they differ based on how the individual was recruited (e.g., an individual with a history of the injury or someone who presents for treatment), time since injury (e.g., immediately vs remotely), or perhaps based on etiology (e.g., motor vehicle accident or sports-related)? For both of these examples above, there are obviously many more moderators that might be differentially related to these broad symptoms domains.

Ecological Validity

Paralleling the field as a whole, few meta-analyses have examined the relationship between neuropsychological variables and ecological validity or aspects of "real world" functioning. Recent exceptions include Kalechstein, Newton, and van Gorp's (2001) meta-analysis on the relationship between neuropsychological test performance and employment status, whereas Reger, Welsh, Watson, Cholerton, Baker, and

Craft (2004) examined the relationship between neuropsychological functioning and driving ability in dementia. Additional meta-analyses in this area would be beneficial, as there has been increasing interest in documenting the ecological validity of neuropsychological tests and their usefulness in making applied, practical decisions.

CONCLUSION

Meta-analysis is a newer method of research synthesis that has become increasingly used in neuropsychology. As several of these have made a significant impact on the science and practice of neuropsychology (e.g., Binder et al., 1997), it behooves clinicians to comprehend the conceptual and statistical basics of such research. Meta-analysis can, in fact, address many questions that are of interest to the neuropsychologist, such as whether cognitive impairment exists in a particular condition or disorder, the nature or signature of such impairment, whether neuropsychological tests are valid or, more to the point, sensitive to such disorder, and what variables are important moderators of the relationship between the variables of interest. With its use of effect sizes, meta-analysis is able to provide quantitative answers to these various questions—traditional methods of research synthesis such as the narrative review do not do so. The quantification of meta-analytic techniques is not only consonant with the manner in which neuropsychological assessment tends to be practiced today, but it also provides a unique method for understanding clinical situations and problems. Moreover, it avoids the now well-known conceptual and statistical problems associated with null hypothesis statistical testing.

REFERENCES

Bangert-Drowns, R. L. (1986). Review of developments in meta-analytic method. *Psychological Bulletin*, *99*, 388–366.

Begg, C. B. (1994). Publication bias. In H. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 399–409). New York: Russell Sage Foundation.

Bezeau, S. & Graves, R. (2001). Statistical power and effect sizes of clinical neuropsychological research. *Journal of Clinical and Experimental Neuropsychology*, *23*, 399–406.

Binder, L. M., Rohling, M. L., & Larrabee, G. J. (1997). A review of mild head trauma: Meta-analytic review of neuropsychological studies. *Journal of Clinical and Experimental Neuropsychology*, *19*, 421–431.

Boone, K. B., Ponton, M. O., Gorsuch, R. L., Gonzalez, J. J., & Miller, B. L. (1998). Factor analysis of four measures of prefrontal lobe functioning. *Archives of Clinical Neuropsychology*, *13*, 585–595.

Christensen, H., Hadzi-Pavlovic, D., & Jacomb, P. (1991). The psychometric differentiation of dementia from normal aging: A meta-analysis. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, *3*, 147–155.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.

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

Demakis, G. J. (2003). A meta-analytic review of the sensitivity of the Wisconsin Card Sorting Test to frontal and lateralized frontal brain damage. *Neuropsychology*, *17*, 255–264.

Demakis, G. J. (2004). Frontal lobe damage and tests of executive processing: A meta-analysis of the category test, stroop test, and trail-making tests. *Journal of Clinical and Experimental Neuropsychology*, *26*, 441–450.

Eysenck, H. J. (1978). An exercise in mega-silliness. *American Psychologist*, *33*, 517.

Feinstein, A. R. (1995). Meta-analysis: Statistical alchemy for the 21st century. *Journal of Clinical Epidemiology*, *48*, 71–79.

Grant, I., Gonzalez, R., Carey, C., Natarajan, L., & Wolfson, T. (2003). Non-acute (residual) neurocognitive effects of cannabis use: A meta-analytic study. *Journal of the International Neuropsychological Society*, *9*, 679–689.

Hedges, L. V. (1994). Fixed effects models. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 285–299). New York: Russell Sage Foundation.

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

Heinrichs, R. & Zakzanis, K. K. (1998). Neurocognitive deficit in schizophrenia: A quantitative review of the evidence. *Neuropsychology*, *12*, 426–445.

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

Jadad, A. R. & McQuay, H. J. (1996). Meta-analyses to evaluate analgesic interventions: A systematic qualitative review of their methodology. *Journal of Epidemiology*, *49*, 235–243.

Johnson-Selfridge, M. & Zalewski, C. (2001). Moderator variables of executive functioning in schizophrenia: Meta-analytic findings. *Schizophrenia Bulletin*, *27*, 305–316.

Kalechstein, A. D., Newton, T. F., & van Gorp, W. G. (2001). Neurocognitive functioning associated with employment status: A quantitative review. *Journal of Clinical and Experimental Neuropsychology*, *25*, 1186–1191.

Kazdin, A. E. (1998). *Research design in clinical psychology* (3rd edition). Needham Heights, MA: Allen and Bacon.

Kindermann, S. & Brown, G. (1997). Depression and memory in the elderly: A meta-analysis. *Journal of Clinical and Experimental Neuropsychology*, *19*, 625–642.

Larrabee, G. J. & Curtiss, G. (1995). Construct validity of various verbal and nonverbal memory tests. *Journal of Clinical and Experimental Neuropsychology*, *17*, 536–547.

Leonberger, F. T., Nicks, S. D., Larrabee, G. J., & Goldfader, P. R. (1992). Facture structure of the Wechsler memory scale-revised within a comprehensive neuropsychological battery. *Neuropsychology*, *6*, 239–249.

Lipsey, M. W. & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications.

Louis, T. A. & Zelterman, D. (1994). Bayesian approaches to research synthesis. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 411–422). New York: Russell Sage Foundation.

Meiran, N. & Jelicic, M. (1995). Implicit memory in Alzheimer's disease: A meta-analysis. *Neuropsychology*, *9*, 291–303.

Meyer-Baron, M., Schaeper, M., & Seeber, A. (2002). A meta-analysis for neurobehavioral results due to occupational mercury exposure. *Archives of Toxicology*, *76*, 127–136.

Morris, S. B. & DeShon, R. P. (2002). Combining effect sizes estimates in meta-analysis with repeated measures and independent-group designs. *Psychological Methods*, *7*, 105–125.

Mountain, M. A. & Snow, W. G. (1993). Wisconsin card sorting test as a measure of frontal pathology: A review. *The Clinical Neuropsychologist*, *7*, 108–118.

Park, N. W. & Ingles, J. L. (2001). Effectiveness of attention rehabilitation after an acquired brain injury: A meta-analysis. *Neuropsychology*, *15*, 199–210.

Raudenbush, S. W. (1994). Random effects models. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 301–321). New York: Russell Sage Foundation.

Reger, M., Welsh, R., Razani, J., Martin, D., & Boone, K. (2002). A meta-analysis of the neuropsychological sequelae of HIV infection. *Journal of the International Neuropsychological Society*, *8*, 410–424.

Reger, M., Welsh, R. K., Watson, G. S., Cholerton, B., Baker, L. D., & Craft, S. (2004). The relationship between neuropsychological functioning and driving ability in dementia: A meta-analysis. *Neuropsychology*, *18*, 85–93.

Rohling, M. R. & Demakis, G. J. (2006). A meta-analysis of the neuropsychological effects of occupational exposure to mercury. *The Clinical Neuropsychologist*.

Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin*, *118*, 183–192.

Rosnow, R. L. & Rosenthal, R. (1989). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, *44*, 1276–1284.

Rossi, J. S. (1990). Statistical power of psychological research: What have we gained in 20 years? *Journal of Consulting and Clinical Psychology*, *58*, 646–656.

Schretlen, D. J. & Shapiro, A. M. (2003). A quantitative review of the effects of traumatic brain injury on cognitive functioning. *International Review of Psychiatry*, *15*, 341–349.

Sharpe, D. (1997). Of apples and oranges, files drawers, and garbage: Why validity issues will not go away. *Clinical Psychology Review*, *17*, 881–901.

Smith, M. L. & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, *32*, 752–760.

Stuss, D. T. & Levine, B. L. (2002). Adult clinical neuropsychology: Lessons from studies of the frontal lobes. *Annual Review of Psychology*, *53*, 401–433.

Vickery, C. D., Berry, D. T. R., Inman, T. H., Harris, H. J., & Orey, S. (2001). Detection of inadequate effort on neuropsychological testing: A meta-analytic review of selected procedures. *Archives of Clinical Neuropsychology*, *16*, 45–73.

Wanous, J. P., Sullivan, S. E., & Malinak, J. (1989). The role of judgment calls in meta-analysis. *Journal of Applied Psychology*, *74*, 259–264.

Wishart, H., & Sharpe, D. (1997). Neuropsycholoigcal ascepts of multiple sclerosis: A quantitative review. *Journal of Clinical and Experimental Neuropsychology*. *19*, 810–824.

Zakzanis, K. K. (1998). Quantitative evidence for neuroanatomic and neuropsychological markers in dementia of the Alzheimer's type. *Journal of Clinical and Experimental Neuropsychology*, *20*, 259–269.

Zakzanis, K. K. (2001). Statistics to tell the truth, the whole truth, and nothing but the truth: formulae, illustrative numerical examples, and heuristic interpretation of effect size analyses for neuropsychological researchers. *Archives of Clinical Neuropsychology*, *16*, 653–667.

Zakzanis, K., Graham, S., & Campbell, Z. (2003). A meta-analysis of structural and functional brain imaging in dementia of the Alzheimer's type: A neuroimaging profile. *Neuropsychology Review*, *13*, 1–18.

Zakzanis, K. K. & Heinrichs, R. W. (1999). Schizophrenia and the frontal brain: A quantitative review. *Journal of the International Neuropsychological Society*, *5*, 556–566.