

Comparability work and the management of difference in research synthesis studies

Margarete Sandelowski^{a,*}, Corrine I. Voils^b, Julie Barroso^c

^a*University of North Carolina at Chapel Hill School of Nursing, USA*

^b*Durham Veterans Affairs Medical Center & Department of Medicine, Duke University Medical Center, USA*

^c*Duke University School of Nursing, USA*

Available online 9 October 2006

Abstract

The new imperative to be more methodologically inclusive has generated a burgeoning interest in synthesizing the findings of qualitative and quantitative studies, or mixed research synthesis. Yet, the very diversity seen to define the mixed research synthesis enterprise is also considered to defy it as it intensifies the problem of comparing the seemingly incomparable to enable the combination of the seemingly uncombinable. We propose here that the research synthesis enterprise, in general, and the mixed research synthesis enterprise, in particular, entail comparability work whereby reviewers impose similarity and difference on the studies to be reviewed. The very study diversity requiring management does not exist a priori but rather is itself an outcome of comparability work already done whereby judgments have been made about what constitutes methodological and topical diversity and uniformity. Conceiving the research synthesis process as defined by comparability work moves the backstage interpretive work of systematic review to center stage and, thereby, sets a new stage for addressing the methodological issues involved. These issues are explored by reference to the synthesis of empirical studies of antiretroviral adherence in HIV-positive women in the US.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Comparability work; Evidence-based practice; Research synthesis; Systematic review; HIV

Introduction

Among the most significant recent developments in the academic and clinical enterprise known as evidence-based healthcare is the call to researchers to be more methodologically inclusive and tolerant of methodological diversity in conducting the systematic reviews at the heart of this practice.

Contributing to this new mandate for inclusion and tolerance are: (a) criticism of evidence-based practice as itself failing to meet its own strict criterion, sustaining an overly narrow view, and even as promoting the loss, of evidence (Timmermans & Berg, 2003; Trinder, 2000; Walker, 2003; White, 2001); (b) criticism of systematic review as reproducing a discredited model of research (Hammersley, 2001); (c) heightened recognition of the inadequacies of randomized controlled trials to address many of the most pervasive health and social problems (Dixon-Woods, Agarwal, Young, Jones, & Sutton, 2004; Mays, Pope, & Popay, 2005); (d) rapid

*Corresponding author.

E-mail addresses: msandelo@email.unc.edu (M. Sandelowski), voils001@mc.duke.edu (C.I. Voils), Julie.Barroso@duke.edu (J. Barroso).

dissemination, and increased calls for the utilization, of qualitative research (Barbour, 2000; Greenhalgh, 2002; Sandelowski, 2004); (e) emergence of mixed methods research as a “third research paradigm” (Johnson & Onwuegbuzie, 2004, p. 14); and (f) postmodern turn toward the examination, accommodation, and celebration of difference (Morris, 2000; Rolfe, 2001).

The convergence of these events has generated a burgeoning interest in synthesizing the findings of methodologically diverse studies (Forbes & Griffiths, 2002; Harden & Thomas, 2005; Hawker, Payne, Kerr, Hardey, & Powell, 2002; Lemmer, Grellier, & Steven, 1999; Popay & Roen, 2003), or mixed research synthesis. Mixed research synthesis studies are systematic reviews of empirical qualitative, quantitative, and mixed methods studies in shared domains of research aimed at aggregating, integrating, or otherwise assembling their findings via the use of qualitative and/or quantitative methods (Sandelowski, Voils, & Barroso, 2006). Yet, the very diversity seen to define the mixed research synthesis enterprise is also considered to defy it. Mixed research synthesis projects dramatize the “heterogeneity” (Deeks, Higgins, & Altman, 2005, Section 8.7) long recognized as the central problem in research synthesis projects; they make even more challenging the task of accommodating the methodological and topical differences that together constitute the unique “personality” (Lipsey & Wilson, 2001) of the studies under review.

In this article, we address this problem of comparing the seemingly incomparable to enable the combination of the seemingly uncombinable. We propose that the research synthesis enterprise is defined by comparability work, which involves finding ways to work with or around study differences. Our purpose here is neither to provide a comprehensive overview of methodological issues per se, nor to advance specific analytical strategies to address them. Rather, our purpose is to move to center stage the backstage interpretive work at the heart of the systematic review process and, thereby, to set a new stage for addressing the methodological issues involved in efforts to synthesize empirical research findings.

Methodological and theoretical location

Our take on method here is as itself an object of inquiry (Law, 2004; Mol, 2002). The immediate impetus for this article were issues raised in the

course of our on-going study directed toward developing methods to combine qualitative and quantitative research findings. The body of literature we chose as the first of several “method cases” to be included in this project is composed of empirical studies of antiretroviral adherence in HIV-positive women of any race/ethnicity, class, or nationality living in the US. The key imperative determining the boundaries we set for the first phase of this project was to have a set of studies methodologically diverse enough to permit, but not so topically diverse as to preclude, fulfilling its methodological aims. Thus far, our study includes 42 reports (35 journal articles, 6 unpublished theses or dissertations, and 1 technical report), retrieved between June, 2005 and January, 2006. Of the 42 reports, 12 are reports of various types of qualitative studies; 3, of intervention studies; 1, of a mixed methods (qualitative phase followed by a pilot intervention) study; and the remainder, of various types of quantitative observational studies.

As we worked our way through several rounds of analysis of the antiretroviral adherence studies selected, we noticed that they were less methodologically diverse than they first appeared. For example, the results of several cross-sectional analyses featuring only one time point were reported from studies designated in the methods section as longitudinal. In addition, the mode of analysis and results of most of the qualitative studies were similar in content, form, and interpretive depth to those presented in several of the quantitative studies. Accordingly, we had fewer longitudinal studies and less varied qualitative studies than we thought we had. In contrast, we noticed that these studies were more topically diverse than we had anticipated. We thought it would be relatively easy to identify antiretroviral adherence studies, yet we found it difficult to distinguish “antiretroviral adherence studies” from studies of such other related topics as antiretroviral medication patterns of use or access. Our concerns that we had less methodological diversity but more topical diversity than we had wanted led us to question our very understanding of study diversity as we found ourselves regularly changing our minds about whether the studies we had selected were methodologically different enough to meet our aims, but topically similar enough to permit their findings to be combined.

Social science perspectives on difference and work

We turned to the social science literature on difference, which helped us to clarify the problem of study diversity and situate it in the larger context of contemporary “ideas of difference” (Hetherington & Munro, 1997). We were attracted to discussions of the politics of difference, especially, of how difference is conceived, articulated, created, sustained, mobilized, and generally managed to accomplish certain ends (Burbules & Rice, 1991; Rosenblum & Travis, 2000). The subject of difference has been addressed in recent social science studies of classification, method, and of evidence-based practice (Bowker & Star, 2000; Law, 2004; Maxwell, 2004; Mol, 2002; Rolfe, 2002). In these literatures, the very idea of difference is troubled as questions are raised concerning what it is, how it is presented, and what difference *difference* makes. *Difference* appears here, not as something that unproblematically defines the relationship between two or more entities, but rather as an ongoing achievement whereby individuals “impose ... similarities and differences” (Steinmetz, 2004, p. 385) to achieve such purposes as order, visibility, and control (Bowker & Star, 2000).

Whereas social science perspectives on difference helped us to clarify the problem of study diversity, the sociology of work (Corbin & Strauss, 1985; Star, 1995) and the “sociology of the invisible” (Star, 1991, p. 267) helped us to re-frame the solution to the problem. In these sociologies, “work is (conceived as) the link between the visible and the invisible” (Star, 1991, p. 265) that allows the exploration of what is “at work” (Mykhalovskiy, McCoy, & Bresalier, 2004, p. 317) but is, nevertheless, often “deleted” (Star, 1995, p. 503) in the research synthesis enterprise.

Our argument

Drawing from these literatures, we propose that all research synthesis projects require the management of difference, or what we refer to here as *comparability work*. The goal of comparability work in the research synthesis process is to make the research findings from different studies in ostensibly the same domain of research comparable enough to be combined. If that is not possible or fails, comparability work involves excluding studies (or selected findings in studies) that resist comparison.

Although they are not conceived as such in the research synthesis or mixed methods literatures, the conversions of qualitative into quantitative data, and of quantitative into qualitative data (Onwuegbuzie & Teddlie, 2003), are instances of comparability work intended to minimize or erase difference: to make the differences deemed to exist between qualitative and quantitative data less visible and even to make these categories themselves less relevant. Similarly, the setting of inclusion and exclusion criteria for the selection of studies to be reviewed, the translation of concepts from two or more qualitative studies into each other (Noblit & Hare, 1988), and the transformation of different statistical expressions of data into various effect size indexes (e.g., *d*, *r*, or odds ratio [Cohen, 1988]) are also instances of comparability work. These indices are “metrics of compatibility and commonness” (Lamont & Molnár, 2002, p. 188) that allow the previously incompatible and uncommon to be compared.

The study diversity that caused us concern was itself an outcome of comparability work that we had already done whereby we had made judgments about what constituted diversity. Indeed, the diversity we sought was not an entity that could be found as it is not a given characteristic of any body of literature, but rather (re)produced in the research synthesis process. Our worry that we had “too little methodological diversity” and “too much topical diversity” was already a judgment of similarity or difference we had imposed while selecting and reviewing studies and that, if we wished, we could choose not to impose.

We now address these new insights in more detail. Although the research synthesis process typically begins with defining the topic of review and, therefore, with resolving issues related to topical diversity, we begin here with a discussion of the methodological diversity seen to define and defy efforts to synthesize findings produced from different kinds of empirical research.

Methodological diversity as object and instance of comparability work

A taken-for-granted assumption in the research synthesis literature is that methodological diversity exists a priori in any body of research under review. The methods classification systems typically appearing in didactic literature on methods draw lines between qualitative and quantitative methods,

between varieties of qualitative methods, and between varieties of quantitative methods on such parameters as philosophical orientation, theoretical foundation, sampling imperative, and approaches to data collection, analysis, and the optimization of validity. These qualitative-versus-quantitative, qualitative-versus-other qualitative, and quantitative-versus-other quantitative research distinctions are viewed as actually shaping the conduct and outcome of inquiry: for example, the way research questions are posed, subjects and events are sampled, and research findings are produced. Indeed, researchers conducting systematic reviews (hereafter referred to as reviewers) are routinely advised to differentiate between the “real” differences found among the groups studied and the “artifactual” differences owing to the way these groups were studied (Glasziou & Sanders, 2002; Hunter & Schmidt, 2004; Song, Sheldon, Sutton, Abrams, & Jones, 2001). A priori methodological distinctions are used in research synthesis projects, among other purposes, to legitimate the inclusion or exclusion of studies, assess adherence to standards of performance for those methods, advance certain types of systematic reviews, or to preserve or secure a privileged place in a hierarchy of methods and evidence (Evans, 2003; Higgins & Green, 2005; Hunter & Schmidt, 2004; Lohr & Carey, 1999; Noblit & Hare, 1988; Ogilvie, Egan, Hamilton, & Petticrew, 2005; Paterson, Thorne, Canam, & Jillings, 2001; Petticrew & Roberts, 2003; Sandelowski & Barroso, 2007).

The qualitative–quantitative distinction

Yet the differences inscribed in methods classification systems and, therefore, assumed to exist between studies designated or designed to be one kind of study versus another (e.g., an observational versus experimental, or qualitative versus quantitative, study) often make no difference to the way studies are actually conducted. Accordingly, the management of methodological diversity in systematic reviews begins with deciding whether and to what extent the studies in the body of literature under review are methodologically similar and different.

A case in point is the distinction typically drawn between “qualitative” and “quantitative” research. Although comparisons between qualitative and quantitative research seem “rhetorically unavoidable” (Becker, 1996, p. 53), in actual practice, the

dividing line between “qualitative” and “quantitative” research is anything but clear as these adjectives are used to designate everything from paradigms of inquiry to techniques for sampling, data collection, and data analysis. When conceived in a “purist” (Johnson & Onwuegbuzie, 2004, p. 14) way—as representing two entirely different species of inquiry—the differences between qualitative and quantitative inquiry appear irreconcilable. The methodological differences between these modes of inquiry are here preserved and even maximized to the point that studies are grouped solely on the basis of being “qualitative” or “quantitative” and then analytically treated differently. In contrast, when conceived in a “compatibilist” (Skrtic, 1990, p. 128) way—as representing simply words versus numbers—the differences appear to be reconcilable by converting the one into the other (De Souza, Gomes, & McCarthy, 2005; Valsiner, 2000). Methodological differences are here minimized or elided as the qualitative–quantitative distinction is deemed irrelevant.

Depending on reviewers’ disciplinary affiliation and philosophical commitments, reports of studies named, framed, and described one way may resemble in content and form reports of studies named, framed, and described in other ways. Conversely, studies named, framed, and described in the same way may not resemble each other at all. One researcher’s “grounded theory” may look to reviewers much like another researcher’s “phenomenology”, and two researchers’ phenomenologies may look to reviewers nothing like phenomenology or each other.

Examples are the “qualitative” reports of studies of antiretroviral adherence in HIV-positive women we selected for review. Most of these reports offered findings in the form of inventories or summaries of the data collected (i.e., lists of facilitators and barriers to adherence, lists of reasons for taking and not taking drugs) that appeared to us to be more similar to the surveys of responses in reports of quantitative studies than to other “qualitative” studies. In another study we had conducted to develop methods to synthesize qualitative findings, we had named these types of reports “topical surveys” (Sandelowski & Barroso, 2007, p. 144) to indicate an in-between or neither-qualitative-nor-quantitative category of inquiry common in the health sciences that share with quantitative surveys an analytic emphasis on condensing the surface informational contents of data via descriptive and statistical summaries.

Accordingly, one management option might be to exclude such studies outright on the grounds that they were not what the authors of the reports of those studies claimed they were, or that they represent weak “qualitative” studies. A second option might be to include these studies but treat their quality as covariates in a posteriori analyses (Cooper, 1998) directed toward ascertaining how each of the studies included in a review contributed to the synthesis of findings produced. By choosing either of these options, difference would be managed by preserving distinctions commonly drawn between qualitative and quantitative research. A third option might be to treat these studies as if they were quantitative surveys (or, at least, treat them in the same way as the quantitative studies they resembled would be treated), which would allow these studies to be included in the review and open up an array of new options for analysis. Here difference would be managed by ignoring the methodological claims made in the reports of these studies and re-classifying them. This option would also ensure that the findings in these studies not be lost to practice for methodological reasons that do not necessarily invalidate findings or make them unusable (Sandelowski & Barroso, 2007).

Real versus nominal methodological diversity

The discrepancy between the way studies are named in reports and the way they were actually conducted (as discernible in, and as judged by reviewers of, those reports) calls into question the difference between, not only, for example, an ethnographic and a not-ethnographic study, and between a good ethnographic study and a bad one (Wolcott, 1990), but also the extent to which methodological distinctions really matter at all on the front lines of studies to the nature and value of the findings produced. As Johnson and Onwuegbuzie (2004) observed, if different methods do not dictate differences in the actual conduct of inquiry, then the differences between them constitute differences without a distinction. Similarly, Becker (1996, p. 57) noted that “philosophical details” often have “little or nothing to do with what researchers actually do”, and Eakin and Mykhalovskiy (2003) proposed that method functions in qualitative research more to stimulate analysis than to determine or constitute findings.

Moreover, few studies of any type meet any one set of performance or quality criteria (Onwuegbuzie

& Daniel, 2003). Studies designed to be grounded theory studies are often actually more equivalent to qualitative descriptive studies, after accounting for the inability to conduct the signature grounded theory processes of theoretical sampling or constant comparison analysis (Charmaz, 2006). Studies designed to be randomized controlled trials are often actually more equivalent to uncontrolled observational studies, after accounting for selection, performance, attrition, and detection biases (Higgins & Green, 2005, Section 6; Wilson & Lipsey, 2001). In the social and behavioral sciences, hardly any studies meet the ideal of the randomized controlled trial, largely because the ethical conduct of research with human subjects and real-world conditions preclude perfect conformity to the “ideal” performance of the randomized controlled trial (Kaptchuk, 2001; Oakley, 1989; Weinstein, 2004). The familiar limitations section appearing in most research reports of studies in the health sciences indicates that methods are easier to honor in the breach than in the observance.

Methods are also not the fixed entities they appear to be in the many quality appraisal tools and checklists promoted (e.g., Sale & Brazil, 2004; West et al., 2002), which have been derived largely from the premise that clearly right and wrong ways exist to execute methods (Chamberlain, 2000). Methods change over time and become what they are in interaction with users. The randomized controlled trial, for example, has undergone many changes to make its findings more credible and its execution more patient-centered and accommodating to real-world conditions (Gross & Fogg, 2001; Kaptchuk, 1998). In actual use, the grounded theory study can acquire a phenomenological cast (Charmaz, 1990), while the descriptive study can be made to appear “more ethnographic” in the reporting of it (Wolcott, 1990). This dynamic relationship between method and user serves ultimately to undermine the distinctions inscribed in methods classification systems.

Further undermining them is that descriptions of method in research reports represent a “reconstructed logic”, as opposed to the “logic-in-use” in the study reported (Kaplan, 1964, p. 8). They may also be instances of “academic posturing” (Wolcott, 1990, p. 45), or efforts to acquire the “epistemological credibility” (Thorne, Kirkham, & MacDonald-Emes, 1997, p. 170) and “rhetorical advantages” (Seale, 2002, p. 659) of naming a study as a certain kind of research. Defining method

(following Wittgenstein) as a “language game”, Gubrium and Holstein (1997, p. 5) observed that “method connotes a manner of viewing and talking about reality as much as it specifies technique and procedure”, and it is this difference between method talk and practice—as opposed to any actual difference between methods per se—that further troubles taken-for-granted ideas about methodological diversity. Research reports are “sites of method talk” (Gubrium & Holstein, 1997, p. 3) at least as much as they are indexes of the studies conducted. Reports are writing practices that make the events constituting the actual studies that were conducted conform to the prescribed forms for reporting them (Bazerman, 1988). This standardization of form conceals the ambiguity, complexity, and sheer messiness of inquiry. Instead of reflecting the actual sequence or content of events, the conventional research report is a “hygiene” move (Law, 2004, p. 9) designed to clean up the mess.

In summary, methodological diversity is not a given characteristic of any set of studies, but rather a judgment reviewers impose on the relationship between studies as discerned from the reports of those studies. Herein lies the comparability work that will influence the inclusion and exclusion of studies and the way the findings in the studies selected for inclusion will be treated.

Topical diversity as object and instance of comparability work

No less an object and instance of comparability work in research synthesis studies than methodological diversity is topical diversity. Systematic reviews typically begin with efforts to define the topic of review with a view toward achieving a topically identical set of studies. Yet, a continuing concern in the systematic review literature is the often-cited “apples and oranges” problem (Deeks et al., 2005, Section 8.1.2; Sharpe, 1997) requiring researchers to decide whether they will treat two entities (e.g., apples and oranges) as one entity (fruit) or preserve the distinctions between them.

A case in point from our antiretroviral adherence studies is the diversity in the entities studied as aspects of “antiretroviral regimen” (e.g., number of pills per dose, number of doses per day, side effects, difficulty swallowing). Comparability work entails deciding whether to treat them as one variable (i.e., antiretroviral regimen) influencing adherence, or to preserve the individuality of each regimen aspect.

The first option would efface what some will regard as regimen features too different to be meaningfully combined, while the second option would leave few findings available for synthesis as relatively few of the adherence studies reviewed (and often only one study) addressed the same regimen aspects. In either option, difference is managed; in the first option, by erasing it as the act of comparing is moved up the continuum of abstraction and in the second, by preserving it as the act of comparing remains at a more empirical level.

According to Glass (2000, p. 6), whose name is virtually synonymous with meta-analysis, to compare apples to apples would not only be “trivial”, but also redundant as nothing but the study of fruit makes sense or is “worthy of true scientists”. Drawing from Nozick’s (1981, p. 29) “closest continuer theory”, Glass concluded that the question “of how two things could be the same ultimately resolves itself into an empirical question” (p. 8) of what researchers conceive of as the important differences. In other words, comparability work is here directed toward deciding what can be seen as similar enough or too different to be combined. In instructional texts on the systematic review of quantitative research, researchers are advised to determine what comparisons should be made and which findings should be used in each comparison. They are also cautioned that such decisions are “inevitably subjective and not amenable to statistical solutions” (Deeks et al., 2005, Section 8.1.2). In his widely used guide to synthesizing quantitative research, Cooper (1998, p. 116) observed that any one “cumulative analysis” should “test the same comparison or estimate the same relationship”. Yet, he also noted that researchers should not combine findings at a level that would elide “distinctions meaningful to the users of the synthesis” (p. 109). As these texts indicate, achieving comparability becomes a matter of technique (e.g., of converting different statistical expressions of data into effect size indexes) only after judgments are made about what comparisons are useful and will appeal to the audiences to which they are directed.

Topical diversity in qualitative versus quantitative studies

In mixed research synthesis studies, comparability work around topical diversity is influenced by divergent views concerning whether qualitative

and quantitative studies can address the same topics. One view is that qualitative and quantitative studies can address the same topic (e.g., participants' "views" of a target event; [Harden et al., 2004](#)), while a contrasting view is that qualitative and quantitative studies are defined, in part, by their addressing different topics: as [Barbour and Barbour \(2003, p. 180\)](#) described it, in comparison to quantitative research, qualitative research "taps into ... a different sort of curiosity".

In the antiretroviral adherence studies in our project, the quantitative findings emphasize numerically measured variables (e.g., CD4 count, viral load, number of pills in the drug regimen) and demographic characteristics (e.g., age, education, injection drug use) as dichotomous or continuous correlates of adherence. In contrast, the qualitative findings emphasize people's experiences with, attitudes toward, and beliefs about antiretroviral therapy. While the quantitative findings in this body of research focus on predictors of extent of adherence or non-adherence, the qualitative findings focus on the reasons for adherence and non-adherence.

Here difference can be managed by treating "predictors" as topically different from "reasons", or conceiving reasons as explanations for predictors. Alternatively, difference can be managed by treating predictors and reasons as equivalent, or conceiving the qualitative findings as more thematically precise versions of the quantitative findings and the quantitative findings as more numerically precise versions of the qualitative findings. In the first instance, difference is imposed whereas in the second, similarity is imposed.

Achieving topical similarity by creating the body of research

Whether the topical differences between qualitative and quantitative studies in a domain of research are preserved or effaced, the management of topical diversity is always complicated by the fact that no two studies of any kind in any body of research deemed to address the same topic ever actually address the same topic, let alone address "it" in the same way. For example, in the body of research addressing HIV-positive women's adherence to antiretroviral therapy are studies directed toward ascertaining what predicts adherence. Yet, these apparently topically similar studies vary widely in the attributes, conditions, events, and other

"factors" studied and the ways in which these factors are conceived, measured, and linked to each other, the way antiretroviral therapy is conceived and measured, the way adherence is conceived and measured, and the persons and sites chosen to study the operation of these factors. *Adherence* in these studies varies widely even in the aspects of number examined, such as percent of time prescription orders are followed, percent of pills taken, and percent of doses consumed over the last day, 2 days, week, or month. In the end, only a few of dozens of studies that researchers will have reviewed in a target domain will have actually addressed the influence of the same set of factors on the same set of other factors in the same way. For example, of the 199 bivariate relationships featured in the quantitative studies we reviewed, 57 were assessed in only 1 study. Moreover, 40 of ostensibly the same relationships (e.g., between education and adherence) were operationalized or analyzed so differently as to resist comparison (e.g., use of different assessment instruments, mean versus dichotomous scoring).

Despite the absolute lack of topical identity between studies in ostensibly the same topical area, the research synthesis enterprise requires reviewers to act as if a designated set of studies is similar enough to treat it as one body of research. Even to talk about *a body of research* is to create an identity between disparate entities. Indeed, reviewers create a body of research for each synthesis project. The very act of deciding what studies to include and exclude is not simply a sampling issue but a form of comparability work whereby reviewers engineer a certain kind of sample. Such engineering efforts are typically directed toward reducing topical diversity in order to have a topically comparable data set for analysis.

Many of the studies we considered for inclusion never actually contained the words *adherence* or *compliance* anywhere in the reports of them and addressed such topics as patterns of access, use, and prescription, and how these patterns correlated with such factors as women's race, class, drug use, psychiatric condition, or CD4 count and viral load. We began our project with the prevailing definition of adherence as something patients do or do not do, namely, follow providers' prescriptions. Yet adherence (i.e., taking medicines as prescribed by providers) does not come into play until a provider actually prescribes medicine for someone, which is, in turn, dependent on the provider seeing that

person as suitable to receive that drug. In addition, adherence does not come into play until individuals are able to fill prescriptions, which is, in turn, dependent on whether they have access to a pharmacy to drop off and pick up the drugs and the means to pay for them. The “arena” (Clarke, 2005) of adherence includes so many more topics than are generally conceived of as constituting the body of adherence research, yet this arena is too topically diverse for any one research synthesis study.

Accordingly, the body of research constituting studies of antiretroviral adherence in HIV-positive women—at any one time for any one research purpose—might include or exclude studies of such topics as: (a) provider practices related to prescribing antiretroviral drugs for HIV-positive women (because these practices determine whether and which women are in a position to adhere, even if these practices were not linked to adherence in the study); (b) factors that facilitate or impede getting prescriptions filled, such as transportation to pharmacies, means to pay (because these factors also determine whether and which women are in a position to adhere, even if these factors were not linked to adherence in the study); (c) side effects of antiretroviral medications (because they may preclude adherence, even if these effects were not linked to adherence in the study); (d) progression of HIV disease (because such disease indexes as CD4 count & viral load also indicate adherence to antiretroviral therapy, even if these indexes were not linked to adherence in the study); (e) provider practices related to selecting and altering specific antiretroviral drugs and drug combinations prescribed (because different drugs and drug regimens will have different effects and, through these effects, influence adherence, even if specific drugs were not linked to adherence in the study); or (f) attitudes toward, beliefs about, or intentions to use antiretroviral drugs (because they can influence adherence practice, even if these factors were not linked to adherence in the study).

As this by no means comprehensive list of contenders for inclusion into the body of antiretroviral-adherence-in-HIV-positive-women research shows, any number or configuration of reports of studies may constitute this body. Indeed, *adherence research* can include studies that never addressed adherence per se, but rather entered the systematic review process because these studies suggested a link between what they did cover and adherence. This

variability in the work object (Law & Singleton, 2005) of systematic review explains why the results of different reviews of ostensibly the same body of research will yield different and even conflicting results (Linde & Willich, 2003) and why it is difficult even to talk about a single body of research or of two systematic reviews of the same body of research.

No one project is likely to contain all of the studies that address the broad terrain of “health work” (Mykhalovskiy et al., 2004, p. 323) in which HIV-positive persons’ medicine-getting and medicine-taking practices are located, that is, the larger “social, discursive, and institutional context(s)” (p. 317) in which HIV-positive persons “do medications” (p. 324). This health work is, in turn, situated in the larger arena of the health work required in other chronic and stigmatizing diseases, which is, in turn, situated in the larger arena of other work that competes with health work, which, is, in turn, situated within still other arenas.

The feasibility of the systematic review enterprise requires cutting a review down to size because, without boundaries, no systematic review is possible. The inclination in systematic review is, therefore, toward exclusion of studies (or findings in studies) to achieve a comparable data set. The systematic review enterprise is less about inclusion, or taking stock (Hunt 1997) of a field of research, and more about exclusion, or finding defensible ways not to take stock of all of it (MacLure, 2005; Torrance, 2004). A systematic review is judged to be credible, in part, to the extent that it makes this “boundary work” (Gieryn, 1983; Lamont & Molnár, 2002) transparent. Reviewers are obliged numerically and narratively to account for every exclusion occurring at each successive stage of the review process, namely, at the search, retrieval, and initial and subsequent data extraction, analysis, and evaluation stages. A report of a study *in* at the search and retrieval phase may be *out* at the data extraction phase as its findings are determined to resist comparison with other findings.

Indeed, the boundary work that defines systematic review is often so exclusionary as to eliminate most of what constitutes the larger arena in which that phenomenon is situated. Reviewers are typically in the strange position of actually synthesizing the findings of only a handful of the reports meeting their initial search criteria. The bias toward exclusion has generated criticisms of the systematic review enterprise as a disciplinary technology

aimed at amassing reasons not to include studies (MacLure, 2005). Yet boundary work is necessary to achieve a manageable (i.e., comparable) data set. In the end, what constitutes a body of research for review is the result of topical differences that reviewers have found ways to “bridge” or “pacify” (Harbers, 2005, p. 578).

Conclusion

We have moved to center stage, and reinterpreted as comparability work, the “judgments, choices, and compromises” (Nurius & Yeaton, 1987) that define the systematic review process, in general, and the research synthesis process, in particular. Comparability work links the procedural transparency, reproducibility, and objectivity said to distinguish systematic from non-systematic reviews (Gough & Elbourne, 2002) with the often “hidden judgments” (Nurius & Yeaton, 1987) upon which the results of systematic reviews inevitably rest. Even systematic reviews conceived as interpretive devices to recast (as opposed to take stock of) a body of research (Livingston, 1999), or to unsettle (as opposed to settle) the nature of or solution to a health or social problem (Eisenhart, 1998), entail comparability work as researchers decide which studies will be reviewed and which findings will be used.

The comparability work that defines the behind-the-scenes work of systematic reviews shows them to be procedurally transparent and reproducible (i.e., objective) only in their encompassing a system of clearly defined tasks (e.g., problem identification, setting of inclusion criteria, selection of search terms and sources, data extraction) to which reviewers can claim adherence and around which they organize their reports (Higgins & Green, 2005). Yet, what is immediately transparent and replicable is only adherence to the tasks and to a style for reporting them, not the enactment of those tasks. Indeed, systematic reviews are reliably unreliable as any one review is a product of the comparability work that defines the unique interaction between reviewers and the body of research they created for review.

The systematic review process, and the mixed research synthesis process, in particular, depend on finding solutions to the problem of difference that will be acceptable to the communities to which they are directed. These communities are themselves differentiated, in part, by their views of study diversity, what study differences matter, and how these differences should be managed. The objectiv-

ity of systematic reviews inescapably rests on making the subjectivity of the process transparent because, in the end, any research synthesis is a product of what gets combined. What gets combined is, in turn, dependent on what gets compared. What gets compared is, in turn, dependent on judgments about what is comparable. What is judged comparable is, in turn, dependent on judgments about similarity and difference. And it is this series of begettings that constitutes the system in systematic review.

Acknowledgements

The study featured here, entitled Integrating Qualitative & Quantitative Research Findings, is funded by the National Institute of Nursing Research, National Institutes of Health, 5R01NR004907. We also acknowledge Career Development Award # MRP 04-216-1 granted to Dr. Voils from the Health Services Research & Development Service of the Department of Veterans Affairs. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs.

References

- Barbour, R. S. (2000). The role of qualitative research in broadening the “evidence base” for clinical practice. *Journal of Evaluation in Clinical Practice*, 6, 155–163.
- Barbour, R. S., & Barbour, M. (2003). Evaluating and synthesizing qualitative research: The need to develop a distinctive approach. *Journal of Evaluation in Clinical Practice*, 9, 179–186.
- Bazerman, C. (1988). *Shaping written knowledge: The genre and activity of the experimental article in science*. Madison: University of Wisconsin Press.
- Becker, H. S. (1996). The epistemology of qualitative research. In R. Jessor, A. Colby, & R. A. Shweder (Eds.), *Ethnography and human development: Context and meaning in social inquiry* (pp. 53–71). Chicago: University of Chicago Press.
- Bowker, G. C., & Star, S. L. (2000). *Sorting things out: Classification and its consequences*. Cambridge, MA: MIT Press.
- Burbules, N. C., & Rice, S. (1991). Dialogue across differences: Continuing the conversation. *Harvard Educational Review*, 61, 393–416.
- Chamberlain, K. (2000). Methodolatry and qualitative health research. *Journal of Health Psychology*, 5, 285–296.
- Charmaz, K. (1990). “Discovering” chronic illness: Using grounded theory. *Social Science and Medicine*, 30, 1161–1172.
- Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. Thousand Oaks, CA: Sage.

- Clarke, A. E. (2005). *Situational analysis: Grounded theory after the postmodern turn*. Thousand Oaks, CA: Sage.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). Hillsdale, NJ: Erlbaum.
- Cooper, H. (1998). *Synthesizing research: A guide for literature reviews* (3rd ed). Thousand Oaks, CA: Sage.
- Corbin, J., & Strauss, A. (1985). Managing chronic illness at home: Three lines of work. *Qualitative Sociology*, 8, 224–247.
- Deeks, J.J., Higgins, J.P.T., & Altman, D.G. (Eds.). (2005). Analyzing and presenting results. In Higgins, J.P.T. & Green, S. (Eds.), *Cochrane handbook for systematic reviews of interventions (version 4.2.5, Section 8)*. Retrieved January 15, 2006 from www.cochrane.org/resources/handbook/hbook.htm.
- De Souza, M. L., Gomes, W. B., & McCarthy, S. (2005). Reversible relationship between quantitative and qualitative data in self-consciousness research: A normative semiotic model for the phenomenological dialogue between data and capta. *Quality and Quantity*, 39, 199–215.
- Dixon-Woods, M., Agarwal, S., Young, B., Jones, D., & Sutton, A. (2004). *Integrative approaches to qualitative and quantitative evidence*. London: Health Development Agency. Retrieved November 13, 2005 from http://www.med.umich.edu/csp/Course%20materials/Summer%202005/Goold_Integrative_Approaches.pdf
- Eakin, J. M., & Mykhalovskiy, E. (2003). Reframing the evaluation of qualitative health research: Reflections on a review of appraisal guidelines on the health sciences. *Journal of Evaluation in Clinical Practice*, 9, 187–194.
- Eisenhart, M. A. (1998). On the subject of interpretive reviews. *Review of Educational Research*, 68, 391–399.
- Evans, D. (2003). Hierarchy of evidence: A framework for ranking evidence evaluating health care interventions. *Journal of Clinical Nursing*, 12, 77–84.
- Forbes, A., & Griffiths, P. (2002). Methodological strategies for the identification and synthesis of “evidence” to support decision-making in relation to complex healthcare systems and practices. *Nursing Inquiry*, 9, 141–155.
- Gieryn, T. F. (1983). Boundary-work and the demarcation of science from non-science: Strains and interests in professional ideologies of scientists. *American Sociological Review*, 48, 781–795.
- Glass, G.V. (2000). Meta-analysis at 25. Retrieved 12/24/2005 from <http://glass.ed.asu.edu/gene/papers/meta25.html>
- Glasziou, P. P., & Sanders, S. L. (2002). Investigating causes of heterogeneity in systematic reviews. *Statistics in Medicine*, 21, 1503–1511.
- Gough, D., & Elbourne, D. (2002). Systematic review synthesis to inform policy, practice and democratic debate. *Social Policy and Society*, 1, 225–236.
- Greenhalgh, T. (2002). Integrating qualitative research into evidence based practice. *Endocrinology and Metabolism Clinics of North America*, 31, 583–601.
- Gross, D., & Fogg, L. (2001). Clinical trials in the 21st century: The case for participant-centered research. *Research in Nursing and Health*, 24, 530–539.
- Gubrium, J. F., & Holstein, J. A. (1997). *The new language of qualitative method*. New York: Oxford University Press.
- Hammersley, M. (2001). On “systematic” reviews of research literatures: A “narrative” response to Evans & Benefield. *British Educational Research Journal*, 27, 543–554.
- Harbers, H. (2005). Science in context: The paradoxes of universalism and liberalism (Book review essay). *Science, Technology, and Human Values*, 30, 575–582.
- Harden, A., Garcia, J., Oliver, S., Rees, R., Shepherd, J., Brunton, G., et al. (2004). Applying systematic review methods to studies of people’s views: An example from public health research. *Journal of Epidemiology and Community Health*, 58, 794–800.
- Harden, A., & Thomas, J. (2005). Methodological issues in combining diverse study types in systematic reviews. *International Journal of Social Research Methodology*, 8, 257–271.
- Hawker, S., Payne, S., Kerr, C., Hardey, M., & Powell, J. (2002). Appraising the evidence: Reviewing the disparate data systematically. *Qualitative Health Research*, 12, 1284–1299.
- Hetherington, K., & Munro, R. (1997). *Ideas of difference: Social spaces and the labor of division*. Oxford, UK: Blackwell.
- Higgins, J.P.T., & Green, S. (Eds.). (2005). *Cochrane handbook for systematic reviews of interventions (4.2.5)*. Retrieved January 15, 2006 from www.cochrane.org/resources/handbook/hbook.htm.
- Hunt, M. (1997). *How science takes stock*. New York: Russell Sage Foundation.
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings* (2nd ed). Thousand Oaks, CA: Sage.
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33, 14–26.
- Kaplan, A. (1964). *The conduct of inquiry: Methodology for behavioral science*. San Francisco: Chandler.
- Kaptchuk, T. J. (1998). Intentional ignorance: A history of blind assessment and placebo controls in medicine. *Bulletin of the History of Medicine*, 72, 389–433.
- Kaptchuk, T. J. (2001). The double-blind, randomized, placebo-controlled trial: Gold standard or golden calf? *Journal of Clinical Epidemiology*, 54, 541–549.
- Lamont, M., & Molnár, V. (2002). The study of boundaries in the social sciences. *Annual Review of Sociology*, 28, 167–195.
- Law, J. (2004). *After method: Mess in social science research*. London: Routledge.
- Law, J., & Singleton, V. (2005). Object lessons. *Organization*, 12, 331–355.
- Lemmer, B., Grellier, R., & Steven, J. (1999). Systematic review of nonrandom and qualitative research literature: Exploring and uncovering an evidence base for health visiting and decision making. *Qualitative Health Research*, 9, 315–328.
- Linde, K., & Willich, S. N. (2003). How objective are systematic reviews? Differences between reviews on complementary medicine. *Journal of the Royal Society of Medicine*, 96, 17–22.
- Lipsey, M. W., & Wilson, D. B. (2001). The way in which intervention studies have “personality” and why it is important to meta-analysis. *Evaluation and the Health Professions*, 24, 236–254.
- Livingston, G. (1999). Beyond watching over established ways: A review as recasting the literature, recasting the lived. *Review of Educational Research*, 69, 9–19.
- Lohr, K. N., & Carey, T. S. (1999). Assessing “best evidence”: Issues in grading the quality of studies for systematic reviews. *Joint Commission Journal on Quality Improvement*, 25, 470–479.

- MacLure, M. (2005). "Clarity bordering on stupidity": Where's the quality in systematic review? *Journal of Education Policy*, 20, 393–416.
- Maxwell, J. A. (2004). Reemergent scientism, postmodernism, and dialogue across differences. *Qualitative Inquiry*, 10, 35–41.
- Mays, N., Pope, C., & Popay, J. (2005). Systematically reviewing qualitative and quantitative evidence to inform management and policy-making in the health field. *Journal of Health Services Research and Policy*, 10(Suppl. 1), 6–20.
- Mol, A. (2002). *The body multiple: Ontology in medical practice*. Durham: Duke University Press.
- Morris, D. B. (2000). How to speak postmodern: Medicine, illness, and cultural change. *Hastings Center Report*, 30, 7–16.
- Mykhalovskiy, E., McCoy, L., & Bresalier, M. (2004). Compliance/adherence, HIV, and the critique of medical power. *Social Theory and Health*, 2, 315–340.
- Noblit, G. W., & Hare, R. D. (1988). *Meta-ethnography: Synthesizing qualitative studies*. Newbury Park, CA: Sage.
- Nozick, R. (1981). *Philosophical explanations*. Cambridge, MA: Belknap Press.
- Nurius, P. S., & Yeaton, W. H. (1987). Research synthesis reviews: An illustrated critique of "hidden" judgments, choices, and compromises. *Clinical Psychology Review*, 7, 695–714.
- Oakley, A. (1989). Who's afraid of the randomized controlled trial? Some dilemmas of the scientific method and "good" research practice. *Women and Health*, 15(4), 25–59.
- Ogilvie, D., Egan, M., Hamilton, V., & Petticrew, M. (2005). Systematic reviews of health effects of social interventions: 2. Best available evidence: How low should you go? *Journal of Epidemiology and Community Health*, 59, 886–892.
- Onwuegbuzie, A. J., & Daniel, L. G. (2003). Typology of analytical and interpretational errors in quantitative and qualitative educational research. *Current Issues in Education*, 6(2). On line at <http://cie.ed.asu/volume6/number2/>
- Onwuegbuzie, A. J., & Teddlie, C. (2003). A framework for analyzing data in mixed methods research. In A. Tashakkorin, & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 351–383). Thousand Oaks, CA: Sage.
- Paterson, B. L., Thorne, S. E., Canam, C., & Jillings, C. (2001). *Meta-study of qualitative health research: A practical guide to meta-analysis and meta-synthesis*. Thousand Oaks, CA: Sage.
- Petticrew, M., & Roberts, H. (2003). Evidence, hierarchies, and typologies: Horses for courses. *Journal of Epidemiology and Community Health*, 57, 527–533.
- Popay, J., & Roen, K. (2003). Synthesis of evidence from research using diverse study designs: A preliminary review of methodological work. Unpublished report for the Social Care Institute of Excellence, Lancaster University, UK.
- Rolfe, G. (2001). Postmodernism for health care workers in 13 easy steps. *Nurse Education Today*, 21, 38–47.
- Rolfe, G. (2002). Faking a difference: Evidence-based nursing and the illusion of diversity. *Nurse Education Today*, 22, 3–12.
- Rosenblum, K. E., & Travis, T.-M. C. (2000). *The meaning of difference: American constructions of race, sex and gender, social class, and sexual orientation* (2nd ed). Boston, MA: McGraw Hill.
- Sale, J. E. M., & Brazil, K. (2004). A strategy to identify critical appraisal criteria for primary mixed-method studies. *Quality and Quantity*, 38, 351–365.
- Sandelowski, M. (2004). Using qualitative research. *Qualitative Health Research*, 14, 1366–1386.
- Sandelowski, M., & Barroso, J. (2007). *Handbook for synthesizing qualitative research*. New York: Springer.
- Sandelowski, M., Voils, C. I., & Barroso, J. (2006). Defining and designing mixed research synthesis studies. *Research in the Schools*, 13, 29–40.
- Seale, C. (2002). Computer-assisted analysis of qualitative interview data. In J. F. Gubrium, & J. A. Holstein (Eds.), *Handbook of interview research* (pp. 651–670). Thousand Oaks, CA: Sage.
- Sharpe, D. (1997). Of apples and oranges, file drawers and garbage: Why validity issues in meta-analysis will not go away. *Clinical Psychology Review*, 17, 881–901.
- Skrtec, T. M. (1990). Social accommodation: Toward a dialogical discourse in educational inquiry. In E. G. Guba (Ed.), *The paradigm dialog* (pp. 125–135). Newbury Park, CA: Sage.
- Song, F., Sheldon, T. A., Sutton, A. J., Abrams, K. R., & Jones, D. R. (2001). Methods for exploring heterogeneity in meta-analysis. *Evaluation and the Health Professions*, 24, 126–151.
- Star, S. L. (1991). The sociology of the invisible: The primacy of work in the writings of Anselm Strauss. In D. R. Maines (Ed.), *Social organization and social process: Essays in honor of Anselm Strauss* (pp. 265–283). New York: Aldine De Gruyten.
- Star, S. L. (1995). Epilogue: Work and practice in social studies of science, medicine, and technology. *Science, Technology, and Human Values*, 20, 501–507.
- Steinmetz, G. (2004). Odious comparisons: Incommensurability, the case study, and "small N's" in sociology. *Sociological Theory*, 22, 371–400.
- Thorne, S., Kirkham, S. R., & MacDonald-Emes, J. (1997). Interpretive description: A noncategorical qualitative alternative for developing nursing knowledge. *Research in Nursing and Health*, 20, 169–177.
- Timmermans, S., & Berg, M. (2003). *The gold standard: The challenge of evidence-based medicine and standardization in health care*. Philadelphia, PA: Temple University Press.
- Torrance, H. (2004). Systematic reviewing—the 'call centre' version of research synthesis. Time for a more flexible approach. *Paper presented to ESRC/RCBN seminar on systematic reviewing*, University of Sheffield. Retrieved May 6, 2006 from, <http://www.esri.mmu.ac.uk/respapers/papers-pdf/seminar-systematicreviewing.pdf>.
- Trinder, L. (2000). A critical appraisal of evidence-based practice. In L. Trinder, & S. Reynolds (Eds.), *Evidence-based practice: A critical appraisal* (pp. 212–241). Oxford, UK: Blackwell Science.
- Valsiner, J. (2000). Data as representations: Contextualizing qualitative and quantitative research strategies. *Social Science Information*, 39, 99–113.
- Walker, K. (2003). Why evidence-based practice now? A polemic. *Nursing Inquiry*, 10, 145–155.
- Weinstein, M. (2004). Randomized design and the myth of certain knowledge: Guinea pig narratives and cultural critique. *Qualitative Inquiry*, 10, 246–260.
- West, S., King, V., Carey, T. S., Lohr, K. N., McKoy, N., Sutton, S. F. et al. (2002). Systems to rate the strength of scientific evidence. Evidence report/technology assessment no. 47,

- AHRQ Pub. No. 02-E016. Rockville, MD: Agency for Healthcare Research and Quality.
- White, D. G. (2001). Evaluating evidence and making judgments of study quality: Loss of evidence and risks to policy and practice decisions. *Critical Public Health*, 11, 1–17.
- Wilson, D. B., & Lipsey, M. W. (2001). The role of method in treatment effectiveness research: Evidence from meta-analysis. *Psychological Methods*, 6, 413–429.
- Wolcott, H. F. (1990). Making a study “more ethnographic”. *Journal of Contemporary Ethnography*, 19, 44–72.