



Published in final edited form as:
Field methods. 2008 ; 20(1): 3–25.

Making Sense of Qualitative and Quantitative Findings in Mixed Research Synthesis Studies

CORRINE I. VOILS,

Durham Veterans Affairs Medical Center and Duke University Medical Center

MARGARETE SANDELOWSKI,

The University of North Carolina at Chapel Hill School of Nursing

JULIE BARROSO, and

Duke University School of Nursing

VICTOR HASSELBLAD

Clinical Research Institute, Duke University Medical Center

Abstract

The synthesis of qualitative and quantitative research findings is increasingly promoted, but many of the conceptual and methodological issues it raises have yet to be fully understood and resolved. In this article, we describe how we handled issues encountered in efforts to synthesize the findings in forty-two reports of studies of antiretroviral adherence in HIV-positive women in the course of an ongoing study to develop methods to synthesize qualitative and quantitative research findings in common domains of health-related research. Working with these reports underscored the importance of looking past method claims and ideals and directly at the findings themselves, differentiating between aggregative syntheses in which findings are assimilated and interpretive syntheses in which they are configured, and understanding the judgments involved in designating relationships between findings as confirmatory, divergent, or complementary.

Keywords

antiretroviral adherence; HIV/AIDS; qualitative research; quantitative research; research synthesis; systematic review; women

Influenced by the turn to evidence-based practice and renewed concerns to enhance the utilization value of academic and clinical research, scholars in the health and social sciences have shown a growing interest in conducting mixed research synthesis studies in which qualitative and quantitative research findings in shared domains of empirical research are integrated (e.g., Harden and Thomas 2005; Sandelowski, Voils, and Barroso 2006). Such studies are increasingly promoted, but many of the conceptual and methodological issues they raise have yet to be fully understood and resolved. In this article we discuss how we managed the issues we encountered while attempting to integrate a set of qualitative and quantitative findings in an ongoing study, the purpose of which is to develop methods to synthesize qualitative and quantitative research findings in common domains of health-related research.

THE MIXED RESEARCH SYNTHESIS PROJECT

We began this method project with studies of antiretroviral adherence conducted with HIV-positive women of any race/ethnicity, class, or nationality living in the United States. These delimitations were set to secure an initial sample methodologically diverse enough to permit but not so topically diverse as to preclude the methodological experimentation at the heart of

the project. Reports of these studies were retrieved using all major channels of communication, primarily forty databases housing citations to literature across the health, behavioral, and social sciences. To accommodate the methodological objectives of the project, we chose a broad and inclusive direction for our synthesis efforts; that is, we were interested in empirical research findings—derived from HIV-positive women themselves—concerning their use of anti-retroviral therapy.

We retrieved forty-two reports (six unpublished master's theses or doctoral dissertations and thirty-six peer-reviewed articles) meeting our search criteria between June 2005 and January 2006. (A list of these reports is available from the authors on request.) Of these forty-two reports, twenty-six are reports of quantitative observational studies, twelve of qualitative studies, three of intervention studies, and one of a mixed methods (qualitative descriptive and pilot intervention) study. From each report, we extracted information on research purpose, design, and methodology. None of these reports were excluded for reasons of quality, as the value of a report for any research synthesis can be determined only while conducting that synthesis (Pawson 2006).

PREPARING THE FINDINGS FOR SYNTHESIS

Because the central objective of our method project was to find ways to synthesize qualitative and quantitative research findings, we separated the qualitative from the quantitative reports. As the differences between qualitative and quantitative research are generally assumed to be the most important obstacles to synthesis (Sandelowski, Voils, and Barroso 2007), we hoped that by treating them separately, we would be able to clarify what distinguishes qualitative from quantitative findings and, therefore, what has to be done to make them combinable.

Preparing the Qualitative Findings

As a result of our initial review of the qualitative reports, we determined that all but one of them offer survey-level findings. Typically derived from individual or focus group interviews and basic content/theme analysis procedures, such findings remain close to data as given by participants in interviews and are located at the low-inference end of a continuum of qualitative data transformation. They are, therefore, not directly amenable to methods of qualitative research synthesis that depend on highly interpreted findings (e.g., qualitative metasynthesis; Sandelowski and Barroso 2007). Indeed, taken as a whole, the findings in the qualitative reports reviewed are comparable in interpretive depth to the descriptive findings in the quantitative reports reviewed, thereby making the line between qualitative and quantitative study less distinct and the entire sample of reports less methodologically diverse as a group than they first appeared.

A key factor glossed in the mixed research synthesis literature is the lack of actual difference between studies presented as qualitative as opposed to quantitative (Sandelowski, Voils, and Barroso 2007). Typically defaulted, too, are idealized and even polemical depictions of qualitative and quantitative research that do not take into account whether the actual findings under review demonstrate the nuanced, penetrating, and context-sensitive interpretations attributed to qualitative research or the mathematical precision, control of bias, and generalizability attributed to quantitative research. Moreover, method claims are too often belied by findings showing that something other than the method claimed was actually used (e.g., a supposedly phenomenological study in which the findings are at a basic descriptive level). The method of research synthesis selected should, therefore, be one that accommodates the actual nature of the findings under review, not the claims made for them.

Finding a suitable synthesis method—To accommodate the descriptive nature of the qualitative findings, we selected qualitative metasummary to synthesize them. Qualitative

metasummary is an aggregative approach to qualitative research synthesis we had previously developed to accommodate primary qualitative survey findings (Sandelowski and Barroso 2007). We use the word *aggregative* to indicate a quantitatively oriented logic for analysis that is largely directed toward identifying those findings that recur most frequently across reports of studies. Although aggregation tends to be depicted as inappropriate, too imitative of quantitative research synthesis, and, generally, wrong for qualitative research findings (e.g., Noblit and Hare 1988; Barbour and Barbour 2003), much qualitative research in the health sciences is at a basic descriptive level, with survey findings that must be pooled before they can be further interpreted. Although informative and, therefore, worthy of inclusion in research synthesis studies, such findings offer no concepts to synthesize, no metaphors to translate, and no coherent lines of argument to align or develop.

Extracting findings—We extracted as findings any researcher interpretation based on data obtained from HIV-positive women pertaining to antiretroviral therapy by separating them from researchers' (1) presentations of data in support of those findings, (2) references to findings from other studies, (3) descriptions of the analytic procedures (e.g., coding schemes) used to produce the findings, and (4) discussions of the significance of their findings. We extracted findings regardless of sample size, as this meets the qualitative research imperative of taking account of all data no matter how idiosyncratic. In addition, most of the qualitative reports offered no information on the numbers of women linked to any finding.

Grouping and abstracting findings—We then grouped findings judged to be topically similar together into seven categories (beliefs/desires, general, provider relations/health services, HIV health status, personal characteristics/responses/experiences, medication regimen, social support/interactions). All findings except five in the “general” domain (e.g., “adherence is a dynamic process”; “nonadherence can be intentional or unintentional”) lent themselves to further grouping into factors favoring adherence or nonadherence. This grouping also reprised the prevailing logic of the findings in the primary reports. We then eliminated redundancies, edited findings to create concise but comprehensive and comprehensible statements of them, and referenced each of these abstracted findings with the report(s) from which it was derived. To optimize the descriptive and interpretive validity (Maxwell 1992) of this process, we worked forward from the list of extracted findings to the abstracted findings, backward from the abstracted findings to the original extracted findings list, and forward again to the abstracted findings.

Calculating effect sizes—To assess the relative magnitude of the abstracted findings, we calculated their frequency effect sizes (Onwuegbuzie 2003). When applied to the synthesis of qualitative research findings, a frequency effect size indicates the number of times an abstracted finding is repeated across reports. With the report as the unit of analysis, frequency effect sizes were computed by taking the number of reports containing a finding (minus any reports derived from a common parent study with a duplication of the same finding) and dividing this number by the total number of reports (minus any reports derived from a common parent study with a duplication of the same finding).

Summarizing the results—From this work, we ascertained that the four most recurrent qualitative findings involve factors favoring nonadherence, including (1) side effects (92%), (2) equivocalness regarding effectiveness (50%), (3) not wanting others to notice the taking of medications (46%), and (4) having regimens difficult to execute in routine daily schedules (46%). A factor favoring adherence—belief in effectiveness—was the next most prevalent (42%). Of a total of sixty-two abstracted findings, thirty-three were unique, derived from only one report each.

Aligning findings addressing the same factors (shown in Table 1), we also determined instances where the same factors operated in divergent ways and when polar opposites or variations of the same factors were addressed. An example of a factor operating in two ways is that having children favored both adherence (when children were viewed as a reason to stay alive and well) and nonadherence (when their care competed with maternal self-care). An example of a polar opposite is that acceptance of HIV favored adherence, whereas denial of HIV favored nonadherence. We were careful not to assume that because one polarity was addressed, its opposite must be true. Accordingly, although acceptance of HIV was identified as a factor favoring adherence, we would not have assumed that denial of it was a factor favoring nonadherence, unless denial was explicitly addressed. Polar opposites were typically addressed in the same report. An example of variations in common factors are contrasting views of ARVs operating in two ways (e.g., favoring adherence as a symbol of hope and survival, favoring nonadherence as a reminder of HIV).

Preparing the Quantitative Findings

We extracted information about every relationship addressed between medication adherence and another variable. As quantitative analyses require fixed operational definitions, we defined medication adherence as the amount of prescribed medication that was consumed, whether assessed by self-report, pill counts, or a medication event monitoring system. We did not include pharmacy refills, as there was no evidence that women received or took the medication. We treated adherence as the dependent variable because that was the intent in the reports reviewed, although the largely atheoretical and correlational nature of the analyses conducted allowed the possibility that adherence was an antecedent or intervening variable.

Grouping relationships and calculating effect sizes—We grouped the variables linked to adherence in the topical domains of demographics, health services, HIV or general health status, medication regimen, mother–child, provider relationships, psychology, social/cultural factors, and substance abuse. For each of these variables, we listed every relationship, with a reference to the report from which it was extracted, how the variable was treated in the analysis (continuous vs. categorical; categories used), and whether the relationship was adjusted or unadjusted. Where possible, we extracted information that could be used to calculate an effect size and then did so for every pairwise comparison. For example, for a main effect of race/ethnicity with three levels (black, Latina, and white), we calculated the effect size for the difference in adherence between black and white, Latina and white, and black and Latina women.

We reverse-scored relationships in which the dependent variable was nonadherence instead of adherence so that the effect size (Cohen's d) always represented the relationship between adherence and the independent variable. A negative d would, thus, indicate a negative relationship (i.e., adherence was lower among group A than group B, or adherence was negatively associated with A), whereas a positive d would indicate a positive relationship (i.e., adherence was greater among group A than group B, or adherence was positively associated with A). We also scored the relationships so that greater numbers would indicate greater value of the independent variable. The direction of scoring was arbitrary; however, it was necessary to score all relationships in the same direction to make the findings comparable and, thus, combinable. After calculating and double-checking the effect sizes for every pairwise relationship, we excluded nonindependent observations so that for each independent variable, no more than one relationship was contributed by a single participant.

Finding a suitable synthesis method—We initially chose meta-analysis to synthesize the findings, as it is the most common method for mathematically summarizing the relationship between independent and dependent variables. Yet a host of problems (further detailed in Voils

et al. 2007) forced us to exclude entire reports or specific effect sizes in reports, which left us too few effect sizes per independent variable to meta-analyze. Accordingly, we next turned to vote counting, in which a significance level is set as a cutoff, and then each relationship is placed into one of three categories: positive (confirming), negative (disconfirming), or no relationship. The category with the greatest number is then assumed to provide the best estimate of the relationship (Bushman 1994).

The disadvantages of vote-counting procedures are that they do not provide an effect size estimate and do not take into account sample size. The advantage is that they allow incorporation of relationships for which there is too little information to calculate an effect size (e.g., only p value is available from the author), and which are characterized by different statistical treatments of independent variables. These two events had compelled us to exclude relationships from the meta-analysis (see Voils et al. 2007). To preserve the advantages of vote counting but address the disadvantage resulting from dependence on sample size, we performed a modified version of a vote count. That is, because p values are so heavily influenced by sample size, we used the effect size to determine whether a finding was positive. Every d above .20, or what Cohen (1988) considers a “small” effect size, was considered a positive result; every d below .20, including negative valence, was considered a negative result.

Vote counting—To perform the vote count, we translated every pairwise comparison into a hypothesis, specifying the direction of the relationship based on the modal response. That is, if four relationships indicated that $A > B$ and one indicated that $A < B$, the $A > B$ hypothesis was chosen. We then tallied the total number of relationships examining that hypothesis (k). Finally, for each hypothesis, we calculated the ratio of positive (hypothesis-confirming) results to k . Using Cohen’s d allowed us to address the issue of p values being influenced to a large extent by sample size. Yet because this method would force us to exclude findings in which there was insufficient information for calculating the effect size, we also calculated the ratio of positive results using $p \leq .05$ to indicate a positive result.

Summarizing the results—A total of 119 hypotheses (for which d values could be calculated) were examined across the twenty-nine reports of quantitative studies. Most hypotheses were in the domain of psychology, which included a host of cognitive, dispositional, behavioral, and mental health variables. All but fifteen of these hypotheses linked one or more independent variables with “adherence” as the dependent variable. The remaining fifteen hypotheses used “intention to adhere” or “difficulty adhering” as the dependent variable. Of the 119 hypotheses examined, 99 had no relationship or only one relationship (using d or p values) contributing to them (including all of the relationships in which the dependent variable was something other than adherence); therefore, they could not be synthesized, as this entails the combination of at least two entities.

Accordingly, as shown in Table 2, only twenty hypotheses were available for synthesis, all with adherence as the dependent variable. Most studied [by $k(d)$ s] among these were the associations between adherence and education [$k(d) = 7$], CD4 counts [$k(d) = 6$], drug use [$k(d) = 5$], depression [$k(d) = 5$], and age [$k(d) = 5$]. Whether ordered by $k(d)$, $k(p)$, or their corresponding ratios, most noteworthy among the hypotheses were the links between lower adherence and being a drug user [ratio (d) = 100%, $k(d) = 5$; ratio (p) = 80%, $k(p) = 5$], and between greater adherence and higher CD4 counts [ratio (d) = 83%, $k(d) = 6$; ratio (p) = 67%; $k(p) = 6$]. We could not further interpret these ratios, even though other researchers have done so by concluding, for example, that a ratio of 50% based on a k of 2 is weaker or less significant than a ratio of 67% based on a k of 6. We decided this was problematic, as each ratio is based on a different denominator (number of relationships). In addition, as shown in Table 2, different conclusions may be reached depending on whether ratio (d) or ratio (p) is interpreted.

SYNTHESIZING THE QUALITATIVE AND QUANTITATIVE FINDINGS

Having summed up the qualitative and quantitative findings to accommodate the distinctive nature of each set of findings and any relevant qualitative and quantitative research imperatives, we sought to find ways to bring them together.

Synthesis via Assimilation or Configuration

Assimilation—Available options for synthesis include assimilation, whereby findings are incorporated into each other, and configuration, whereby findings are arranged into a theoretical model, narrative line of argument, or other coherent form (Sandelowski, Voils, and Barroso 2006). Assimilation is possible when findings are viewed as confirming each other or converging in the same direction, thereby permitting a conclusion to be drawn regarding their relative magnitude. Assimilation is most closely aligned with aggregative approaches to qualitative and quantitative research synthesis (e.g., qualitative metasummary, meta-analysis, vote counting) in which findings viewed as repetitive are pooled to yield one finding signifying more or less evidence for its existence than other pooled findings.

Configuration—In contrast, configuration is the option when findings are viewed as complementing as opposed to confirming each other. Here findings cannot be merged, but they can be “meshed” (Mason 2006:20), as they are seen to explain or extend each other or otherwise to contribute to an arrangement deemed by reviewers to confer order on those findings (e.g., a conceptual model or map, a meta-narrative). Configuration is most closely aligned with interpretive approaches to research synthesis in which findings are used to generate new or modify existing theoretical or narrative renderings of the target events under review. Indeed, even though it is not referred to as such, configuration appears to be the prevailing mode of synthesis advanced for integrating qualitative findings and increasingly advanced for integrating qualitative and quantitative findings (e.g., Harden et al. 2004; Dixon-Woods et al. 2006).

Assimilation versus configuration—Assimilations of findings are databased syntheses that reflect common aspects of a target phenomenon that has actually been addressed. Databased syntheses signify findings anchored in or demonstrably grounded in or supported by primary research findings and an obligation not to veer far from those findings. In contrast to assimilation, configuration allows findings to be used as jumping-off points or linked to each other in ways never addressed in the primary reports that yielded them. Because they allow reviewers more free rein in selecting the findings that will be used and in how they will be put together, configurations of findings are best seen as data-generated or data-inspired syntheses composed of new ways to see a target event that have yet to be studied as configured. Configurations recast or may even unsettle a domain of study (Eisenhart 1998; Livingston 1999), but they also require further study to establish their value in guiding future research or practice in that domain.

Challenges of assimilation and configuration—Assimilation was a challenge because of the different units of analysis we had to use to synthesize the qualitative and quantitative findings, respectively, to accommodate their distinctive natures and distinctive qualitative and quantitative research imperatives. Assimilation requires that a common metric or language be found to combine findings. The unit of analysis used to combine the qualitative findings was the number of reports in which a unique finding appeared regardless of sample size. This choice met the qualitative research imperative of accounting for all data and accommodated the lack of information in the qualitative reports concerning the numbers of women expressing any topic or theme. In contrast, the unit of analysis used with the quantitative findings was the number of relationships (themselves based on sample size) contributing to a hypothesis for

which a *d* or *p* value could be computed. This choice met the requirements of vote counting and accommodated the diversity of relationships actually addressed in the quantitative reports.

Moreover, whereas the quantitative effect sizes were based on the average effect across all subjects and thereby ignored whether a relationship existed for any one subject, the effect sizes of the qualitative findings were based on the presence of findings across reports, even if present in only one report and based on only one subject. Although qualitative findings may show the same variable working in opposing ways under different conditions (e.g., having children favoring adherence when children are viewed as a reason to live and nonadherence when child care competes with maternal self-care), the same variable operating in opposing ways in quantitative studies will yield a statistically nonsignificant main effect. (Moderators would illuminate the conditions under which a variable operates, but few authors examined moderators, and no two authors examined the same moderating relationships.) Configuration was also a challenge, as the range and diversity of findings and relative lack of any theoretical or interpretive staging for them in the primary reports meant that we might have to move outside these reports and further into our imaginations to lend coherence to them.

Comparing the qualitative and quantitative findings at the group level, we found the topical domains they address to be grossly similar, although this is likely a result, in part, of our working on these data sets both concurrently and sequentially and communicating with each other in weekly research team meetings (as opposed to having different members of the research team working completely separately from each other). These topical domains are also grossly comparable to those recurrently appearing in the antiretroviral adherence reviews or state-of-the-science literatures (e.g., Fogarty et al. 2002; Castro 2005) with which we are familiar. Yet the entities studied within each of these topical domains are highly diverse, with the quantitative findings emphasizing largely demographic and clinical factors and the qualitative findings emphasizing women's beliefs and relationships. This within-topic diversity is also a recurrent feature of antiretroviral adherence literature in which, for example, diverse items are all studied as indicators of medication regimen.

Further complicating either the assimilation or configuration of findings were primary qualitative findings that do not permit clear lines to be drawn between adherence (behavior or what women actually did) and their beliefs or intentions. As a consequence, the synthesized qualitative findings do not distinguish between them either and, therefore, include them all. In contrast, the primary quantitative findings do allow adherence behavior to be distinguished, but there were no hypotheses addressing intentions to adhere or perceived difficulty adhering to which at least two relationships contributed. As a result, the synthesized quantitative findings include only findings on adherence behavior.

Taking the Synthesized Quantitative Findings as the Comparative Reference Point

With these group level distinctions in mind, we took the synthesized quantitative findings as the comparative reference point to ascertain the relationship of the synthesized qualitative findings to them. This approach was virtually impossible, as the quantitative findings were composed of explicit comparisons (e.g., A higher or lower than B vis-à-vis adherence), whereas the qualitative findings offered no such comparisons or implied them. The qualitative findings could, therefore, not be translated into the terms of the quantitative findings. This key difference between the synthesized qualitative and quantitative findings is captured in Sivesind's (1999) contrast between the "single dimensionality," ideally characterizing quantitative research or its orientation toward ascertaining differences between specified groups on a selected and relatively small number of specified variables, and the "singularity," ideally characterizing qualitative research or its orientation toward delineating the complex particularities of the case.

The only exceptions are two general findings from the Schrimshaw, Siegel, and Lekas (2005) report that (1) women in the pre-HAART era (1994–1996) were more likely to report negative attitudes and intolerable side effects and less likely to report perceived benefits than women in the HAART era (2000–2003) and that (2) African American women in the pre-HAART and HAART eras were more likely to report negative attitudes and less likely to perceive benefits than Puerto Rican or white women in both eras. The former finding could not be linked to any other qualitative or quantitative finding. The latter appears to complement the qualitative finding that African American women tended to view antiretroviral medications as racist or genocidal and the quantitative findings linking black women to lower adherence than Latina or white women.

Taking the Synthesized Qualitative Findings as the Comparative Reference Point

Taking the synthesized qualitative findings as the comparative reference point to ascertain the relationship of the quantitative findings to them was more useful. Because most of the qualitative findings feature factors favoring adherence or nonadherence, we could translate the quantitative findings into those terms for further comparison and combination. Although the quantitative hypothesis “adherence was greater among women with more education” is equivalent to the hypothesis “adherence was lower among women with less education” (because the quantitative effect sizes were based on correlations having a range of values that allow the relationship between adherence and another variable to be stated both ways), for consistency, greater than hypotheses were translated to factors favoring adherence, whereas lower/less than hypotheses were translated to factors favoring nonadherence.

Table 3 arranges only those qualitative and qualitatively translated findings that appeared to us to be related in some way: confirming, diverging from, and/or complementing (extending or explaining) each other. We judged a significant proportion of findings to be unrelated to any other finding, reducing the number of findings that could be included in the synthesis of qualitative and quantitative research findings. The interpretive complexity of discerning the relationships between findings is a process that tends to be glossed in the research synthesis literature. Whether and how two or more entities are seen to be related are themselves judgments derived from the clinical and research knowledge and inclinations to discerning sameness and difference that reviewers bring into the synthesis enterprise. Moreover, such judgments are complicated by ambiguities in the findings themselves.

The best illustration of this is the Set G findings #16–21 shown in Table 3. Addressing the link between health status and symptoms and adherence, no comfortable conclusion can be drawn from these findings, as they can be variously read as confirming, contradicting, complementing, or even having no relationship to each other. For example, a higher CD4 count suggests better health and fewer symptoms and might, therefore, lead reviewers to assume the existence of a complementary relationship favoring adherence between having no symptoms/feeling healthy and higher CD4 counts and a divergent relationship with higher detectable viral load. Yet it is not necessarily the case clinically that a person with a higher CD4 count will have no symptoms of HIV or that a higher detectable viral load will yield any symptoms at all. Moreover, the quantitative reports do not specify whether it is the CD4 count per se that favors adherence or whether women’s understanding that a high CD4 count means that they are doing well that favors adherence. Given the correlational nature of the quantitative findings, the direction of causation is also uncertain; a high CD4 count could be the outcome (rather than antecedent) of adherence.

A contradiction exists in this set between the qualitative findings indicating that both feeling healthy and feeling sick favor adherence and that feeling healthy favors both adherence and nonadherence. Such apparent contradictions might be explained away (i.e., turn out not to be contradictions) if the findings themselves address varying conditions under which they might

diverge. Indeed, this was the case with the qualitative findings #5–6 in Set C that having children favored adherence and nonadherence under different conditions (i.e., when they were seen as a reason to live and when their care competed with self-care). Without such attention to variations, however, no satisfying conclusion can be drawn about this relationship.

In short, the relationships shown in Table 3 constitute our best guesstimate based on what made sense to us. We only surmised, for example, that the findings #24–27 in Set I were in a complementary relationship, as prior research and common sense suggest that having more education and the wisdom garnered from age and the experience of having taken antiretroviral medication for a while could explain why knowledge of HIV and these medications might favor adherence.

Such sense making may draw from the reports in a targeted body of research or from theoretical or empirical knowledge outside that body of research. For example, three of the general qualitative findings—addressing the dynamism, intentionality, and narratives of adherence—could be used as conceptual or metaphoric devices for configuring both sets of findings. Used this way, they lend a general coherence to the diverse and diversely operating factors found to favor adherence and nonadherence. That is, these factors could be brought together by the qualitative finding that adherence is a dynamic process whereby women alternated between intentional and unintentional adherence and nonadherence. This configuration aligns also with a view of adherence proposed outside the forty-two reports featured here as involving dose-by-dose decisions (Wilson, Hutchinson, and Holzemer 2002) made on a case-by-case basis or of adherence as “episodic” (Ryan and Wagner 2003:796). Indeed, the dose-by-dose/case-by-case concept could be imported to conceptually synthesize (i.e., configure) the findings. An alternative configuration, derived from the only qualitative report featuring a narrative as opposed to survey treatment of data (Sankar et al. 2002), is that women’s responses concerning adherence can be understood not as indexes of actual experiences (e.g., behaviors, beliefs) but rather as discourses in which different sources of authority (e.g., provider, family) and different moral accounts prevail to justify varying patterns of adherence.

CONCLUSION

Although all of the reports we reviewed address ostensibly the same topic (antiretroviral adherence), our effort to put the findings in these reports together revealed that few of them deal with the same topic in the same way. This is likely why reviews and state-of-the-science literature on antiretroviral adherence frequently end with the conclusion that although much has been studied, little is actually generalizably true concerning antiretroviral adherence that can serve as the basis for interventions to improve it (e.g., Ammassari et al. 2002; Reynolds 2004). The resistance of these findings to synthesis is also a function of the relatively atheoretical, acultural, and ahistorical way in which adherence was (and continues to be) studied (e.g., Bresalier et al. 2002; Broyles, Colbert, and Erlen 2005). Most of the quantitative reports address an assortment of variables without benefit of a priori theoretical staging. Most of the qualitative reports feature an assortment of responses with virtually no a posteriori interpretive staging. Both sets of findings were, thus, composed of isolated data bits resisting efforts to make them cohere. Neither the qualitative nor quantitative set of findings could enhance the meaning or significance of the other, as neither delivered on the distinctive advantages ideally attributed to qualitative research (e.g., nuanced description, penetrating interpretation) or quantitative research (e.g., precise conceptualization and measurement, sophisticated statistical analyses), which are advanced as the central reason for mixing methods and findings (Sandelowski 2004).

Working with these forty-two reports affirmed to us the importance of taking a qualitative in addition to a quantitative approach toward the synthesis of qualitative and quantitative findings.

The qualitative approach was more useful for the findings we had, as the synthesized quantitative findings could be translated into the terms of the qualitative findings. The prevailing solutions advanced for combining qualitative and quantitative data have entailed largely “quantitizing” qualitative data or using qualitative findings to accessorize quantitative findings. Recent scholarship, however, has called for more “qualitizing.” Mason (2006:10) urged more “qualitative thinking,” and Howe (2004:42) promoted “mixed-methods interpretivism” over the prevailing “mixed-methods experimentalism” as ways to transcend the conventional qualitative/quantitative divide and to offset the priority usually given to quantitative thinking and methods.

Yet what qualitative versus quantitative thinking and methods mean for the mixed research synthesis enterprise remains unclear, in part, because this binary reproduces false distinctions. As we have already suggested, many of the reports designated as qualitative call into question what exactly defines a qualitative study. Moreover, qualitative thinking may, in fact, include the use of quantitative techniques to integrate qualitative findings, as meaning is inescapably numbered. The patterns and themes that regularly appear in qualitative research reports imply a perceived recurrence of things judged to be the same (Fredericks and Miller 1997). At the same time, qualitative thinking always entails a tilt toward inclusion, as the qualitative imperative is to try to make sense of all data, no matter how disparate, inconsistent, or ambiguous they may first appear to be. Qualitative meta-summary is an example of qualitative thinking implemented via counting that, nevertheless, meets the qualitative research imperatives concerning the preservation of data in all of its ambiguity. In quantitative research, the tilt is to exclusion, as quantitative synthesis is constrained by the mandate to meet statistical assumptions.

Another manifestation of the qualitative/quantitative divide is the bias against aggregation for the synthesis of qualitative findings because it is deemed to betray quantitative (or positivist) thinking. Yet seeking to ascertain which findings are supported by “a preponderance of evidence” (Thorne et al. 2004:1362) is not a betrayal of qualitative research; nor is the use of numbers what distinguishes qualitative from quantitative research. Moreover, the fact remains that much qualitative research in the health sciences yields low-inference survey findings that simply do not lend themselves to interpretive synthesis without prior aggregation. Indeed, arguably, some form of counting seems to be a requirement even for interpretive synthesis, if only to ensure that all findings are taken into account and that the patterns and themes of which the synthesis is composed are justified (Fredericks and Miller 1997).

In the end, research synthesis projects are best designed by reflexive doing—a principle of qualitative research design—as opposed to being done by fixed a priori design. Reviewers simply cannot know in advance what any set of findings will allow or enter synthesis projects already committed to a synthesis approach. Just as the value of a report for any research synthesis can be determined only in the course of conducting that synthesis, so the value of any synthesis method can be determined only by looking past idealized notions of what findings qualitative and quantitative methods ought to produce and directly at the findings themselves.

Acknowledgements

The study featured here, titled “Integrating Qualitative & Quantitative Research Findings,” is funded by the National Institute of Nursing Research, National Institutes of Health, 5R01NR004907, June 3, 2005–March 31, 2010. We also acknowledge Career Development Award no. MRP 04-216-1 granted to the first author from the Health Services Research and Development Service of the Department of Veterans Affairs. The views in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs.

References

- Ammassari A, Trota MP, Murri R, Castelli F, Narciso P, Noto P, Vecchiet J, D'Arminio-Montforte A, Wu AW, Antinori A. ADIoNA Study Group. Correlates and predictors of adherence to highly active antiretroviral therapy: Overview of published literature. *JAIDS—Journal of Acquired Immune Deficiency Syndromes* 2002;31(Supplement 3):S123–27.
- Barbour RS, Barbour M. Evaluating and synthesizing qualitative research: The need to develop a distinctive approach. *Journal of Evaluation in Clinical Practice* 2003;9(2):179–86. [PubMed: 12787181]
- Bresalier, M.; Gillis, L.; McClure, C.; McCoy, L.; Mykhalovskiy, E.; Taylor, D.; Webber, M. Making care visible: Antiretroviral therapy and the health work of people living with HIV/AIDS. 2002 [accessed on October 1, 2006]. <http://cbr.cbrc.net/files/1052421030/makingcarevisible.pdf#search=%22making%20care%20visible%22>
- Broyles LM, Colbert AM, Erlen JA. Medication practice and feminist thought: A theoretical and ethical response to adherence in HIV/AIDS. *Bioethics* 2005;19(4):362–78. [PubMed: 16222853]
- Bushman, B. Vote-counting procedures in meta-analysis. In: Cooper, H.; Hedges, LV., editors. *The handbook of research synthesis*. New York: Russell Sage; 1994. p. 193-214.
- Castro, A. Adherence to antiretroviral therapy: Merging the clinical and social course of AIDS; *PloS Medicine*. 2005 [accessed on January 8, 2007]. p. 1217\–21. Open access journal, <http://medicine.plosjournals.org/perlserv/?request=get-document&doi=10.1371%2Fjournal.pmed.0020338>
- Cohen, J. *Statistical power analysis for the behavioral sciences*. 2. Hillsdale, NJ: Lawrence Erlbaum; 1988.
- Dixon-Woods, M.; Cavers, D.; Agarwal, S.; Annandale, E.; Arthur, A.; Harvey, J.; Hsu, R.; Katbamna, S.; Olsen, R.; Smith, L.; Riley, R.; Sutton, AJ. Conducting a critical interpretive synthesis of the literature on access to healthcare by vulnerable groups. *BMC Medical Research Methodology*. 2006 [accessed on December 27, 2006]. Open access journal, <http://www.biomedcentral.com/1471-2288/6/35>
- Eisenhart MA. On the subject of interpretive reviews. *Review of Educational Research* 1998;68(4):391–99.
- Fogarty L, Roter D, Larson S, Burke J, Gillespie J, Levy R. Patient adherence to HIV medication regimens: A review of published and abstract reports. *Patient Education and Counseling* 2002;46(2): 93–108. [PubMed: 11867239]
- Fredericks M, Miller SI. Some brief notes on the “unfinished business” of qualitative inquiry. *Quality & Quantity* 1997;31(1):1–13.
- Harden A, Garcia J, Oliver S, Rees R, Shepherd J, Brunton G, Oakley A. Applying systematic review methods to studies of people’s views: An example from public health research. *Journal of Epidemiology and Community Health* 2004;58:794–800. [PubMed: 15310807]
- Harden A, Thomas J. Methodological issues in combining diverse study types in systematic reviews. *International Journal of Social Research Methodology* 2005;8(3):257–71.
- Howe KR. A critique of experimentalism. *Qualitative Inquiry* 2004;10(4):42–61.
- Livingston G. Beyond watching over established ways: A review as recasting the literature, recasting the lived. *Review of Educational Research* 1999;69(1):9–19.
- Mason J. Mixing methods in a qualitatively driven way. *Qualitative Research* 2006;6(1):9–25.
- Maxwell JA. Understanding and validity in qualitative research. *Harvard Educational Review* 1992;62(3):279–300.
- Noblit, GW.; Hare, RD. *Meta-ethnography: Synthesizing qualitative studies*. Newbury Park, CA: Sage; 1988.
- Onwuegbuzie AJ. Effect sizes in qualitative research: A prolegomenon. *Quality & Quantity* 2003;37(4): 393–409.
- Pawson R. Digging for nuggets: How “bad” research can yield “good” evidence. *International Journal of Social Research Methodology* 2006;9(2):127–42.
- Reynolds NR. Adherence to antiretroviral therapies: State of the science. *Current HIV Research* 2004;2(3):207–14. [PubMed: 15279584]

- Ryan GW, Wagner GJ. Pill taking “routinization”: A critical factor to understanding episodic medication adherence. *AIDS Care* 2003;15(6):795–806. [PubMed: 14617501]
- Sandelowski M. Using qualitative research. *Qualitative Health Research* 2004;14(10):1366–86. [PubMed: 15538005]
- Sandelowski M.; Barroso, J. Handbook for synthesizing qualitative research. New York: Springer; 2007.
- Sandelowski M, Voils CI, Barroso J. Defining and designing mixed research synthesis studies. *Research in the Schools* 2006;13(1):29–40.
- Sandelowski M, Voils CI, Barroso J. Comparability work and the management of difference in research synthesis studies. *Social Science & Medicine* 2007;64(1):236–47. [PubMed: 17029691]
- Sankar A, Luborsky M, Schuman P, Roberts G. Adherence discourse among African-American women taking HAART. *AIDS Care* 2002;14(2):203–18. [PubMed: 11940279]
- Schrimshaw EW, Siegel K, Lekas H-M. Changes in attitudes toward antiviral medication: A comparison of women living with HIV/AIDS in the pre-HAART and HAART eras. *AIDS and Behavior* 2005;9(3):267–79. [PubMed: 16088368]
- Sivesind KH. Structured, qualitative comparison: Between singularity and single-dimensionality. *Quality & Quantity* 1999;33(4):361–80.
- Thorne S, Jensen L, Kearney MH, Noblit G, Sandelowski M. Qualitative meta-synthesis: Reflections on methodological orientation and ideological agenda. *Qualitative Health Research* 2004;14(10):1342–65. [PubMed: 15538004]
- Voils CI, Barroso J, Hasselblad V, Sandelowski M. In or out? Methodological considerations for including and excluding findings from a meta-analysis of predictors of anti-retroviral adherence in HIV-positive women. *Journal of Advanced Nursing* 2007;59(2):163–77. [PubMed: 17543011]
- Wilson HS, Hutchinson SA, Holzemer WL. Reconciling incompatibilities: A grounded theory of HIV medication adherence and symptom management. *Qualitative Health Research* 2002;12(10):1309–22. [PubMed: 12474905]

Biographies

CORRINE I. VOILS, PhD, is an assistant professor of medicine in the Center for Health Services Research in Primary Care at the Durham Veterans Affairs Medical Center and in the Department of Medicine at Duke University. Her primary research interest is understanding and improving treatment adherence. Recent publications include “Five-Year Trajectories of Social Networks and Social Support in Older Adults with Major Depression” (with J. C. Allaire et al., International Psychogeriatrics, forthcoming) and “In or out? Methodological Considerations for Including and Excluding Findings from a Meta-Analysis of Predictors of Antiretroviral Adherence in HIV-Positive Women” (with J. Barroso, V. Hasselblad, and M. Sandelowski, Journal of Advanced Nursing, 2007).

MARGARETE SANDELOWSKI, PhD, RN, FAAN, is the Cary C. Boshamer Professor at the University of North Carolina at Chapel Hill School of Nursing. Her research interests include gender, technology, health care, and methods development. Recent publications include Handbook for Synthesizing Qualitative Research (with J. Barroso, Springer, 2007), “Comparability Work and the Management of Difference in Research Synthesis Studies” (with C. I. Voils and J. Barroso, Social Science & Medicine, 2007), and “‘Meta-jeopardy’: The Crisis of Representation in Qualitative Metasynthesis” (Nursing Outlook, 2006).

JULIE BARROSO, PhD, ANP, APRN, BC, is an associate professor at the Duke University School of Nursing. Her research interests include qualitative methods, HIV-related fatigue, and issues with HIV-positive women. Recent publications include “From Synthesis to Script: Transforming Qualitative Research Findings for Use in Practice” (with M. Sandelowski et al., Qualitative Health Research, 2006), “Research Results Have Expiration Dates: Ensuring Timely Systematic Reviews” (with M. Sandelowski and C. I. Voils, Journal of Evaluation in Clinical Practice, 2006), and “Using Qualitative Metasummary to Synthesize Qualitative and

Quantitative Descriptive Research Findings” (with M. Sandelowski and C. I. Voils, Research in Nursing and Health, 2007).

VICTOR HASSELBLAD, PhD, is a professor of biostatistics in the Department of Biostatistics & Bioinformatics at the Duke Clinical Research Institute of Duke University. His research interests include meta-analysis, clinical trial methods, noninferiority, power, distribution fitting, and dose-response analysis. Recent publications include “Prediction of Rehospitalization and Death in Severe Heart Failure by Physicians and Nurses of the ESCAPE Trial” (with L. M. Yamokoski et al., Journal of Cardiac Failure, 2007), “The Cobalt Chromium Stent with Antiproliferative for Restenosis II (COSTAR II) Trial Study Design: Advancing the Active-Control Evaluation of Second-Generation Drug-Eluting Stents” (with T. Y. Wang et al., American Heart Journal, 2007), and “Discussion of: Statistical and Regulatory Issues with the Application of Propensity Score Analysis to Nonrandomized Medical Device Clinical Studies” (with Y. Lokhnygina and M. W. Krucoff, Journal of Biopharmaceutical Studies, 2007).

TABLE 1
 Alignment of Synthesized Qualitative Findings Addressing the Same Factors ($n = 13$)

Favoring Adherence	K	ES %	Favoring Nonadherence	k	ES %
Acceptance of being HIV-positive	1	8	Personal Characteristics/Responses/Previous Experience		
Knowledge/understanding of HIV and ARVs	2	15	Denial of, ambivalence about, HIV; negative emotions/emotional trauma associated with chronicity and uncertainty of HIV	5	38
Active stance toward adherence (e.g., agency located in self, personal responsibility)	1	8	Lack of knowledge/understanding of HIV and ARVs; information overload	1	8
Seeing a difference, other HIV-positive people live long and well on ARVs	2	15	Passive stance toward adherence (e.g., agency located in clinic, God)	1	8
Having, or seeing others have, HIV-negative baby on ARVs	2	15	Not seeing, not noticing any difference with ARVs	1	8
			Having or seeing others have HIV-negative baby without ARVs	1	8
			Taking alternatives to ARVs (e.g., vitamins, positive thinking)	1	8
			Priority given to individual/body wisdom	1	8
			Beliefs, Intentions, Desires		
Belief in effectiveness, advantages, and lack of harm of ARVs	5	42	Equivocal about effectiveness, and short- and long-term toxicity of ARVs (to self or fetus/infant)	6	50
Viewing ARVs as symbol of hope and survival, way to live longer	4	31	Viewing ARVs as reminder of HIV/deviant status	3	23
Viewing ARVs as way to control one's fate/disease	1	8	Viewing ARVs as racist, genocidal (African American women)	2	15
Obligation to baby	1	8	Belief that AZT is unnecessary if mother is healthy	1	8
Having confidence that can take ARVs as prescribed	1	8	Not having confidence that can take ARVs as prescribed	1	8
			Social Support/Interactions		
Absence of reference to or concern about stigma (except in case of children)	1	8	Not wanting others to notice or know HIV status	6	46
Having children (to live for them)	4	31	Having children (caring/competes with medicine taking)	2	15
Family/friends have positive view, are supportive, remind to take	4	31	Family/friends have negative view, are not supportive, cause distress	3	23
Largely scientific authority in support of adherence	1	8	Largely personal, popular, and/or conflicting authority in support of selective adherence	1	8
			Provider Relations, Health Services		
Supportive, trustworthy, accessible, or demonstrably caring MD/provider	5	38	Unsupportive, untrustworthy, inaccessible, or demonstrably uncaring MD/provider	3	23
			HIV/General Health Status		
Having no symptoms of HIV, feeling healthy	3	23	Having no symptoms of HIV, feeling healthy or better	5	38
Feeling sick, having symptoms	2	15	Feeling sick	1	8
			Medication Regimen		
Having no side effects or manageable side effects of ARV	2	15	Side effects of ARV	12	92
Having less complex regimen or one that allows integration into routine schedule	4	31	Having ARV regimen that is difficult to execute in routine daily schedule (e.g., forget, asleep)	6	46
			Having ARV regimen that is difficult to execute in nonroutine schedule (e.g., vacation, away from home)	1	8
			Pills hard to take, too many	4	31

NOTE: ES = effect size; ARV = antiretroviral; AZT = zidovudine (an antiviral drug that inhibits replication of some retroviruses and is used to treat AIDS)

TABLE 2
 Synthesis of Quantitative Findings with at Least Two Contributing Relationships Ordered by $k(d)$ ($n = 29$)

<i>Hypothesis: Adherence Was ...</i>	K(d)	Ratio (d)	k(p)	Ratio (p)	Domain
greater among women with greater education	7	14	7	14	D
greater among women with greater CD4 counts	6	83	6	67	HHS
lower among illegal drug users than nonusers	5	100	5	80	P
lower among women with depression	5	40	5	40	P
greater among older than younger women	5	20	6	17	D
lower among women drinking alcohol	4	100	4	25	SA
greater among women with greater/detectable viral load	4	50	4	50	HHS
greater among women who were optimistic	3	33	3	33	P
lower among women on PI/HAART	3	33	3	33	MR
greater among women with less frequent dosing	2	100	2	100	MR
lower among black than white women	2	50	3	0	D
greater among women taking ARV >2 years	2	50	2	50	MR
greater among women with at least one child	2	50	2	0	MC
greater among women who had opportunistic Infections	2	50	2	0	HHS
greater among employed than unemployed Women	2	50	3	33	D
greater among black than Latina women	2	0	2	0	D
lower among Latina than white women	2	0	2	0	D
greater among women with anxiety	2	0	2	0	P
greater among women with insurance	1	0	2	0	D
greater among women with partner or married	1	0	2	0	P

NOTE: $k(d)$ = # of relationships contributing to hypothesis; ratio (d) = corresponding frequency effect size or # of reports supporting the hypothesis at $d > .20/k(d)$; $k(p)$ = # of relationships contributing to hypothesis for which there was information about the p value, even if just stating it was not significant. This differs from $k(d)$ when there was no further information available other than stating a relationship was not significant; ratio (p) = corresponding frequency effect size or # of reports supporting the hypothesis at $p < .05/k(p)$. PI/HAART = protease inhibitor/highly active antiretroviral therapy; ARV = antiretroviral; D = demographics; HHS = HIV health status; MR = medication regimen; P = psychological; SA = substance abuse; MC = mother-child.

TABLE 3
Alignment of Synthesized Qualitative and Quantitative Findings Confirming (CONF), Complementing (COMP), or Diverging from (DIV) Each Other

Qualitative Findings	Quantitative Findings	Quantitative Findings
<i>Favoring Adherence (FES, # reports)</i>	<i>Favoring Adherence (ratio (d), # relationships)</i>	<i>Favoring Nonadherence (FES, # reports)</i>
<p>Set A</p> <p>1. Belief in effectiveness, advantages, lack of harm of ARVs (42%, $k = 5$) (CONF #2)</p> <p>Set B</p> <p>3. Supportive, trustworthy, & demonstrably caring provider (38%, $k = 5$) (CONF #5)</p> <p>Set C</p> <p>5. Having children (desire to live for them; 31%, $k = 4$) (DIV #6)</p> <p>Set D</p> <p>7. Family/friends have positive view, are supportive, remind to take (31%, $k = 4$) (CONF #8)</p> <p>Set E</p> <p>9. Having less complex regimen, allows integration into routine schedule (31%, $k = 4$) (CONF #11, COMP #10, 12)</p> <p>Set F</p> <p>13. Viewing ARVs as symbol of hope and survival, way to live longer (31%, $k = 4$) (COMP #14, 15)</p> <p>Set G</p> <p>16. Having no symptoms of HIV, feeling healthy (23%, $n = 3$) (DIV #17, 18, 21; COMP #19)</p> <p>17. Feeling sick, having symptoms (15%, $k = 2$) (DIV #16, 19, 21; COMP #18)</p> <p>Set H</p> <p>22. Having no side effects or manageable side effects of ARV (15%, $k = 2$) (CONF #23)</p> <p>Set I</p> <p>24. Knowledge/understanding of HIV and ARVs (15%, $k = 2$) (COMP #25, 26, 27)</p> <p>Set J</p> <p>Set K</p>	<p>10. Less frequent dosing (100%; $k = 2$) (COMP #9, 11, 12, 13)</p> <p>14. Optimism (33%, $k = 3$) (COMP #13)</p> <p>18. Opportunistic infections (50%, $k = 2$) (COMP #17; DIV #16, 21)</p> <p>19. Higher CD4 counts (83%, $k = 6$) (COMP #16, 21; DIV #17, 20)</p> <p>20. Higher detectable viral load (50%, $k = 4$) (DIV #19)</p> <p>25. More education (14%, $k = 7$) (COMP #24)</p> <p>26. Older (20%, $k = 5$) (COMP #24)</p> <p>27. Taking ARVs > 2 years (50%, $k = 2$) (COMP #24)</p>	<p>2. Equivocal about effectiveness, toxicity of ARVs (50%, $k = 6$) (CONF #1)</p> <p>4. Unsupportive, untrustworthy, inaccessible, or demonstrably uncaring provider (23%, $k = 3$) (CONF #3)</p> <p>6. Having children (caregiving competes with self-care; (15%, $k = 2$) (DIV #5)</p> <p>8. Family/friends have negative view, not supportive, cause distress (23%, $k = 3$) (CONF #7)</p> <p>11. Having ARV regimen difficult to execute in routine daily schedule (46%, $k = 6$) (CONF #9, COMP #10, 12)</p> <p>12. Pills hard to take, too many (31%, $k = 4$) (COMP #9, 10, 11)</p> <p>15. Viewing ARVs as reminder of HIV/deviant status (23%, $k = 3$) (COMP #13)</p> <p>21. Having no symptoms of HIV, feeling healthy or better (38%, $k = 5$) (DIV #16, 17, 18; COMP #19)</p> <p>23. Side effects of ARVs (92%, $k = 12$) (CONF #22)</p> <p>28. Viewing ARVs as racist, genocidal (African American women; 15%, $k = 2$) (COMP #29)</p> <p>30. Negative emotions/trauma associated with HIV (38%, $k = 5$) (CONF #31)</p> <p>29. Black (vs. white; 50%, $k = 2$) (COMP #28)</p> <p>31. Depression (40%, $k = 5$) (CONF #30)</p>

NOTE: ARV = antiretroviral; FES = frequency effect size