Published in final edited form as:

Qual Quant. 2011 April ; 45(3): 653–669. doi:10.1007/s11135-010-9375-z.

Bayesian data augmentation methods for the synthesis of qualitative and quantitative research findings

Jamie L. Crandell,

Department of Biostatistics, University of North Carolina at Chapel Hill, #7460 Carrington Hall, Chapel Hill, NC 27599, USA; School of Nursing, University of North Carolina at Chapel Hill, #7460 Carrington Hall, Chapel Hill, NC 27599, USA

Corrine I. Voils,

Health Services Research & Development Service, Durham Veterans Affairs Medical Center & Duke University Medical Center, Durham, NC, USA

Yunkyung Chang, and

School of Nursing, University of North Carolina at Chapel Hill, #7460 Carrington Hall, Chapel Hill, NC 27599, USA

Margarete Sandelowski

School of Nursing, University of North Carolina at Chapel Hill, #7460 Carrington Hall, Chapel Hill, NC 27599, USA

Abstract

The possible utility of Bayesian methods for the synthesis of qualitative and quantitative research has been repeatedly suggested but insufficiently investigated. In this project, we developed and used a Bayesian method for synthesis, with the goal of identifying factors that influence adherence to HIV medication regimens. We investigated the effect of 10 factors on adherence. Recognizing that not all factors were examined in all studies, we considered standard methods for dealing with missing data and chose a Bayesian data augmentation method. We were able to summarize, rank, and compare the effects of each of the 10 factors on medication adherence. This is a promising methodological development in the synthesis of qualitative and quantitative research.

Keywords

Meta-analysis; Meta-synthesis; Synthesis; Cross-design synthesis; Bayesian data augmentation; Missing data; Gibbs sampling

1 Introduction

Among the challenges to conducting comprehensive systematic reviews of research in the health, behavioral, and social sciences is how best to synthesize the findings from methodologically diverse studies. A spate of literature over the past decade has been focused on describing the issues raised in, proposing methods for, and reporting the results of reviews directed toward integrating qualitative and quantitative research findings (e.g., Dixon-Woods et al. 2005; Harden and Thomas 2005; Pope et al. 2007; Voils et al. 2009). The possibility of using Bayesian methods to integrate qualitative and quantitative research

is mentioned repeatedly in this literature; however, there remains only one published example of the actual use of a Bayesian approach to synthesize qualitative and quantitative evidence (Roberts et al. 2002). Recognizing the need for further research in this area, we described a second Bayesian method for synthesis in a paper currently under review (Voils et al. 2009). In this article, we describe a third method and compare it to these other two Bayesian approaches.

1.1 Orientations to incorporating qualitative research into systematic reviews

What constitutes a synthesis of research, and even qualitative as opposed to quantitative research, is still the subject of methodological debate, with proponents of qualitative research most concerned about the co-optation of qualitative research imperatives to quantitative research standards, and proponents of quantitative research most concerned about maintaining those standards of empirical research (e.g., Barbour and Barbour 2003; Hammersley 2002; Sandelowski et al. 2007). A comprehensive review of these debates is beyond the scope of this article. Accordingly, for our purposes here, we define research synthesis as the actual fusion of qualitative with quantitative evidence—the assimilation of one into the other—as opposed to including or comparing qualitative and quantitative research findings in one systematic review. Although we find the continued binary distinction between qualitative and quantitative research regrettable, we use these terms here as a shorthand communication of the difference between studies directed toward narrative understandings, and range of possibilities, of a target phenomenon (qualitative), and those directed toward ascertaining prevalence and probable influence of one or more pre-specified variables on one or more other pre-specified variables signifying that phenomenon (quantitative).

With these definitions and distinctions in mind, qualitative research can be said to be incorporated into systematic reviews at three different points: before any synthesis of information, as part of the synthesized information, or as a framework in which to interpret synthesized information. Incorporating qualitative evidence before the synthesis involves using it to inform the process, for example, to pose an appropriate research question for a systematic review of the quantitative literature and/or to determine elements of the design to answer the research question, such as properties of the sample(s) and the variables of interest. This reflects a still common understanding of qualitative research as precursor to quantitative research whereby the use of qualitative evidence is limited to generating, but not answering, research questions. Incorporating qualitative evidence at the end of a review, to interpret quantitative results, also places limits on its contribution to the results themselves. If researchers wish to treat qualitative evidence as data along with quantitative evidence, rather than as before or after accessories, the qualitative evidence should be allowed to contribute directly to—that is, to shape—the results of the systematic review.

One way to do this involves integrating the findings from qualitative and quantitative studies separately and then using the results of these two separate analyses to inform the aggregated evidence. This can be done with existing methods if the final synthesis of results is done non-mathematically (see Goldsmith et al. 2007).

A mathematical synthesis of qualitative and quantitative evidence has been performed by Roberts et al. (2002), who described a Bayesian approach they used to identify factors affecting the uptake of childhood immunization. They used evidence from the qualitative studies as prior information, allowing it to inform about which factors might be related to the uptake of immunization in children. Findings from the qualitative studies were tabulated and combined with their own expert opinion to estimate the prior probability that each factor was related to receiving immunization. They then used evidence from the quantitative studies on the same topic to update these prior probabilities, yielding posterior probabilities

that each factor was associated with the decision to immunize. In short, Roberts et al. incorporated information from the qualitative studies at the start of their analysis. They synthesized the qualitative information to get prior probabilities of the relationships and then updated these with evidence from the quantitative studies.

In response to the Roberts et al. (2002) effort, we (Voils et al. 2009) observed that the approach they used did not treat the qualitative evidence with as much weight as the quantitative evidence, and proposed that it be incorporated into the same stage of the analysis as the quantitative evidence to allow both to equally inform about the results. The main challenge in incorporating qualitative and quantitative data at the same stage of the analysis is that, in order for the two types of information to be incorporated in the same way, they must be transformed so that they can be combined.

In Voils et al. (2009), we used qualitative and quantitative observational studies on antiretroviral adherence to determine whether a more complex medication regimen was associated with increased adherence. To make the two types of evidence compatible, the findings in the qualitative studies were quantified using a set of strategies, described in Chang et al. (2009), for inferring the frequencies of findings from reports of qualitative descriptive studies that did not offer this information in numbers. Information about finding frequencies was extracted from the qualitative reports by transforming verbal counts (e.g., *few*, *many*) into numbers. The finding frequencies were designated as ranges because the actual numbers were not available, which is often the case in qualitative research reports (although not necessarily a deficiency in qualitative reports (see Chang et al. (2009))). These frequency ranges were synthesized in the same way as the findings frequencies from the quantitative studies to estimate the probability that, for a given research participant, regimen complexity was associated with increased adherence. To study the feasibility of synthesis under these conditions, the data were synthesized separately for the qualitative and quantitative studies. The two study types yielded different estimates of the probability of an association. We hypothesized that this was due either to the method used to extract findings frequencies, or to differences between subject-level probabilities of association in qualitative and quantitative studies. Yet, this approach was explicit in its attempt to give the same treatment to the frequency data extracted from both the qualitative and quantitative studies.

In the analysis featured in this article, our goal was to perform a different type of Bayesian synthesis that still allowed for the same treatment of quantitative and qualitative reports. In Voils et al. (2009), we made the two types of evidence compatible by extracting subject-level quantitative information from qualitative reports, effectively quantizing the qualitative evidence. In the current project, motivated in part by the lack of availability of subject-level information in the qualitative reports, we chose instead to look at all reports on the study level. We translated the numerical results of the quantitative papers into themes, categorizing each report according to the presence/absence of themes. Once the themes were extracted from the qualitative and quantitative reports, a Bayesian data augmentation method was used to summarize the findings. We now describe this Bayesian synthesis of qualitative and quantitative research findings.

2 Methods

2.1 Collecting research reports

The research reports used in the analysis featured here were collected as part of an ongoing methodological study directed toward developing methods to synthesize qualitative and quantitative research findings. The sample of reports included 12 reports of qualitative descriptive studies and 15 reports of quantitative observational studies with finding related to factors facilitating or hindering antiretroviral adherence. (Further information about the

larger methodological study, including search criteria, are available from the last author on request.)

2.2 Data abstraction

Examining the list of variables in both studies, we found that many variables were similar, but not identical. This is typically the case in systematic reviews; it is usually necessary to combine factors seen to be similar and, therefore, treatable as the same because otherwise, there would be few findings available to synthesize, no matter what the method of synthesis used. We grouped similar variables together; for example, the variables feeling ill, side effects, and weight gain were all considered to be one factor called “side effects.” We ended up with 12 factors potentially related to adherence to study.

2.3 Choosing the form of the data

We treated each research report as a case (Gerring 2007), coding each according to its conclusions about the relationship between each of the factors and adherence. For the quantitative studies, we used Cohen’s d (the standardized difference between means, a commonly used effect size index) and applied criteria such that if $d \geq .20$, it was coded as 1; $-.20 < d < .20$, as .50; and $d \leq -.20$, as 0. For the qualitative studies, if a variable was related to adherence, it was coded as 1, if related to both adherence and nonadherence, or neither, it was coded as .5, and, if related to nonadherence, it was coded as 0. For both types of studies, if a report did not address a certain variable, the cell was left blank, as we were careful not to assume a lack of effect of a variable that was not examined at all. An important difference between qualitative and quantitative research is that in quantitative research, an a priori decision is made concerning what variables will be studied. In contrast, in qualitative research, whether a variable is addressed depends on what happens in the study, for example, on how questions were asked, whether participants brought it up, or on whether researchers chose to feature it.

We first created a data matrix (Table 1) summarizing the findings of all of the reports, with each column corresponding to one of the selected variables and each row to a single report. Entries were made in the matrix whenever a factor was reported as promoting adherence (1), having no effect on adherence (0.5), or promoting nonadherence (0). Table 1 contains only 10 factors. We deleted two of the originally selected factors (substance abuse and good relationship with healthcare provider) because there was no variability among study findings: all studies that addressed these variables found that a good relationship with the healthcare provider was associated with adherence and that substance abuse was associated with nonadherence. The methods we describe are not applicable to factors with no variability, because they rely on the estimation of a variance. The deletion of these factors also led to the removal of two reports because they offered no results on any of the remaining 10 factors.

Returning to Table 1, if each row is treated as an observation and each column as a variable, the cells corresponding to the qualitative and quantitative studies for each of the 10 factors can be treated as a dataset. Because any one report addressed only a small subset of the variables studied across reports, a large proportion of the cells in the dataset remain empty.

2.4 A naive analysis

As shown in Table 1, a large proportion (75%) of cells contain no data. A very basic (naive) analysis of these data could ignore the missing values, and then take the average of the non-missing values in each column to estimate the strength of the relationship between each variable and adherence or nonadherence to an antiretroviral regimen. For each factor, the estimate describing the strength of association is the average of the non-missing cells in the

column. A 95% confidence interval for each mean could also be constructed. This procedure ignores reports that did not address the factor, which means that the number of reports on which an estimate is based will vary from factor to factor. For example, for the association between feeling ill and adherence/nonadherence, the estimate is $(1 + .5 + .5)/12 = .17$. These naive estimates are shown in Table 2. As we noted previously, a value of zero in the table means that a factor is associated with nonadherence to the regimen, a value of .5 means there is no association between the factor and adherence, and a value of 1 means that a factor is associated with adherence. Under the assumption of normality, the 95% confidence intervals will contain the “true” mean value for that variable 95% of the time. If a confidence interval contains .5, there is not enough evidence categorically to state an association with either adherence or nonadherence.

Because the number of observations used to construct each confidence interval is small, the appropriateness of the confidence intervals is based on the unlikely assumption that the data are normally distributed. The normal confidence interval makes no allowance for the fact that the value must range from 0 to 1, so the confidence intervals are truncated to allow no values outside this range; they are reported in the fourth column of Table 2 (i.e., a value <0 becomes 0, and a value >1 becomes 1).

This analysis allows us to draw a conclusion about the relationship between adherence and each variable. For example, the confidence interval for the association between side effects/feeling ill and adherence is (0, .35). This interval does not contain .5, so we would conclude that side effects/feeling ill is associated with nonadherence. A belief that antiretroviral medications might do harm is also associated with nonadherence, with an even narrower confidence interval of (0, .25). In addition to the small sample sizes making the confidence intervals unreliable, they also give us limited power to draw conclusions about the relative strengths of these associations (i.e., examine whether a belief that antiretroviral medications might do harm is a stronger predictor of nonadherence than side effects/feeling ill). To do this, we would have had to do a *t*-test (or nonparametric test) to compare means, and the sample sizes are simply too small for powerful tests. In order to make adequate use of the small amount of available information, we chose to use a more complex statistical procedure.

2.5 A more rigorous analysis

2.5.1 Background—The naive analysis ignores cells in which the data were missing, using standard methods of estimation to summarize the results in the non-missing cells. This gives unsatisfactory results, as the resulting confidence intervals are based on very small samples and the normality assumption, which is inappropriate.

Having recognized that our main problem was that we had only a small number of observed cells in our dataset, we turned our attention first to standard methods for dealing with missing data. This dataset—as we surmise will be the case with most research synthesis datasets—has more missing data than a typical dataset from a primary research study; therefore, the assumptions of these methods and their applicability to the current dataset must be carefully considered. Broadly defined, there are two common methods for dealing with missing data: complete case analysis and Bayesian data augmentation.

Complete case analysis: Complete case analysis involves deleting all cases that are incomplete, then analyzing the rest (listwise deletion). Complete case analysis in research synthesis projects is problematic in two ways. The first is enough to thwart most attempts to use complete case analysis; that is, rarely will complete observations be available in a research synthesis dataset. A complete case occurs only when investigators report results

about all of the variables of interest, therefore, retaining only complete cases in our dataset would have eliminated all of the observations.

A second problem is that complete case analysis can bias results if missing values are not missing completely at random (MCAR). In this case, MCAR is true if the reports that did not have findings about a given variable are in no way systematically different from the reports that did have findings about the variable. This situation would arise if the variables to include in each study were chosen completely at random, which is unlikely because large, comprehensive studies probably have different aims and procedures than smaller, more focused studies. Qualitative studies have more inclusive, open-ended aims and research questions that function like a “compass,” while quantitative studies have more specific, fixed aims that function like an “anchor” (Eakin and Mykhalovskiy 2003, p. 190).

Bayesian data augmentation: Whenever analyzing a dataset using complete case analysis is deemed unacceptable due to excessive loss of data (i.e., few or no complete cases) and/or possible violation of the MCAR assumption, there are several methods that can be employed (Little and Rubin 2002). The most widely-used of these methods relies on employing the available data to impute the values of the missing data with Bayesian methods, then analyzing the data as if it were complete. The use of observed values to inform about plausible missing values is called *imputation* and requires the assumption that all missing values exist, but are not observed. Therefore, each row of the dataset must be conceived as a vector, of which only a few elements are visible. Moreover, it must make sense for a value to appear in each cell for which data are imputed.

The Bayesian data augmentation method that we used (Gelman et al. 2004) is relatively straightforward to implement and is appropriate for any amount of missing data. This method assumes that each row in the table, if completely observed, could be considered a random vector drawn from some type of multivariate distribution. In reality, only a few elements from each draw are observed (i.e., each vector is only partially observed). The data augmentation procedure involves using the observed data to estimate the unknown parameters of the multivariate distribution, then drawing from the distribution to estimate the missing values themselves. The estimated missing values are then used to fill in the empty cells in the dataset (the name, *data augmentation*, comes from the fact that observed data are augmented with imputed data), and the parameters of the distribution are re-estimated. This procedure is repeated until a large number of estimates have been obtained. The estimates are then summarized to describe the parameters of the multivariate distribution from which each row was drawn.

2.5.2 Details of the rigorous analysis—Each row of the table can be thought of as an independent draw from a multivariate normal distribution, although, due to the nature of the data, we truncated the multivariate normal distribution so that each value was constrained to fall between 0 and 1. By truncating this way, we obtained a distribution that covers our desired range, while still retaining some of the mathematically advantageous properties of the normal distribution.

To glean as much information as we could out of this sparse (i.e., containing a high proportion of missing values) dataset, we used the observed values to impute missing values. The algorithm begins by looking at all of the non-missing values to learn about the relationships among variables. Then it uses those relationships to impute missing values from the observed values in each row. This is often called “borrowing” information across observations (e.g., Berry 2006 in a clinical trials setting; Choi et al. 2007 in a meta-analysis). The data in Table 1 offer a simple non-mathematical example. Roberts and Mann (2000) reported that both threats to confidentiality and side effects were associated with

nonadherence (i.e., have a value of 0). Schrimshaw et al. (2005) also found a link between side effects and nonadherence, but reported nothing about threats to confidentiality. If we borrowed information from the Roberts and Mann study, we might surmise that there would also have been a relationship between threats to confidentiality and nonadherence in the Schrimshaw et al. study. The actual imputation is much more complex than this, simultaneously borrowing information from all variables at once. Such imputation allows for relationships among variables, but does not force relationships among variables where the observed data support none. The imputed values, unlike the real data, are not restricted to 0, 0.5, and 1, but can take any value ranging from 0 to 1. The imputed values can, therefore, reflect varying degrees of support for a relationship with adherence/nonadherence, or the lack of any relationship. For example, an imputed value of 0.75 would support equally the idea of no relationship (.5) and a relationship with adherence (1).

Another useful feature of Bayesian thinking is that it allows parameters to have statistical distributions. We were interested in the posterior distributions of the mean vector and the covariance matrix of the multivariate normal distribution from which each (only partially observed) row was assumed to have arisen. We used an iterative procedure, known as a Gibbs Sampling algorithm (Casella and George 1992), which draws samples from the posterior distributions of the mean and covariance. When a large number of samples has been obtained, they can be summarized (means, percentiles) to describe the posterior distribution of the parameters.

Let $\mathbf{Y}$ be an $n \times p$ matrix of data with missing values and let $\mathbf{y}_i$ denote the $p \times 1$ vector of elements in the i th row of the data matrix. $\mathbf{y}_i$ can be subdivided into missing (unknown) values, $\mathbf{y}_{i,mis}$ (a $p_{i,mis} \times 1$ vector) and observed values, $\mathbf{y}_{i,obs}$ (a $p_{i,obs} \times 1$ vector). Note that $p_{i,mis} + p_{i,obs} = p$ for $i = 1, \dots, n$. We assume the following likelihood for each of the n rows of the data matrix:

$$\mathbf{y}_i | TN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}), i=1, \dots, n$$

where n is the number of reports. $TN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes a p -dimensional multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, truncated so that every element falls in $[0,1]$. Because not all the elements of $\mathbf{y}_i$ are observed directly, we have chosen to impute the missing values at each iteration by using the conditional distribution of the missing elements on the observed elements. Conditional distributions of the multivariate truncated normal are multivariate truncated normal (Horrace 2005) so

$$\mathbf{y}_{i,mis} | \mathbf{y}_{i,obs} \sim TN_{12}(\boldsymbol{\mu}_{i,mis} + \boldsymbol{\Sigma}_{i,mo} \boldsymbol{\Sigma}_{i,obs}^{-1} (\mathbf{y}_{i,obs} - \boldsymbol{\mu}_{i,obs}), \boldsymbol{\Sigma}_{i,mis} - \boldsymbol{\Sigma}_{i,mo} \boldsymbol{\Sigma}_{i,obs}^{-1} \boldsymbol{\Sigma}_{i,om}), \quad i=1, \dots, n$$

where the parameters are the same as if it were a non-truncated conditional multivariate normal distribution: $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ have been partitioned to correspond to the means and covariances of the observed and missing values. For the i th observation, $\boldsymbol{\mu}_{i,obs}$ and $\boldsymbol{\Sigma}_{i,obs}$ are the mean and covariance of the observed values, and $\boldsymbol{\mu}_{i,mis}$ and $\boldsymbol{\Sigma}_{i,mis}$ are the mean and covariance of the missing values. The covariance between the elements of $\mathbf{y}_{i,mis}$ and $\mathbf{y}_{i,obs}$ is the $p_{i,mis} \times p_{i,obs}$ matrix $\boldsymbol{\Sigma}_{i,mo}$, which is equal to $\boldsymbol{\Sigma}_{i,om}^T$.

The prior structure was developed by putting conjugate priors on $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$:

$$\begin{aligned} \boldsymbol{\mu} | & \quad \Sigma N_p(\boldsymbol{\mu}_0, k_0 * \boldsymbol{\Sigma}) \\ \boldsymbol{\Sigma} | & \quad \text{Inv-Wishart}(v, \boldsymbol{\Sigma}_0 / (v - p - 1)) \end{aligned}$$

Hyperparameters were chosen to carefully balance objectivity and the constraints of the current problem. Because the posterior distribution of μ should have negligible mass outside $[0,1]$, values of ν that gave samples in this range were preferred. Jeffrey's joint prior, a standard non-informative prior, was rejected for violation of this criterion. The prior on μ was chosen so that it had a mean of 0.5 (i.e. no effect) and was relatively flat when elements of μ were between 0 and 1. This was accomplished by letting $\mu_0 = .5 * \mathbf{1}_p$, where $\mathbf{1}_p$ denotes a $p \times 1$ vector of ones and letting $\kappa_0 = 10$. Various choices for the degrees of freedom, ν , of the prior distribution of Σ were examined. A value of $\nu = 100$ was chosen, as it was determined that this value led to samples in the desired range. Much higher values of ν tended to suppress any correlation between factors, and much lower values tended to overstate the extent of that correlation. Σ_0 was taken to be the diagonal matrix containing the variances of the columns of $\mathbf{Y}$.

Due to the conjugate prior structure, the full conditional posterior distributions of μ and Σ were multivariate normal and inverse-Wishart respectively (see Gelman et al. 2004, pp. 87–88 for details). The Gibbs sampling algorithm proceeded by iteratively sampling the unobserved elements of the data matrix, $\mathbf{y}_{i,mis}$ for $i = 1, \dots, n$, from (1), the mean vector, μ , from its full conditional posterior distribution and the covariance matrix, Σ , from its full conditional posterior distribution.

The MCMC algorithm was run in Matlab 7.6.0. Ten million samples were taken, and every hundredth was retained for analysis, for a total of 100,000 samples. The first 10,000 were discarded as a burn-in, leaving 90,000 for analysis. Convergence and adequate mixing of the Gibbs Sampler were observed on traceplots of all parameters.

3 Results

Point estimates and credible intervals (the Bayesian analog to confidence intervals) were constructed by summarizing the 90,000 samples. The point estimate of each mean is the average over all the samples, and can be used to describe the strength of the relationship between each factor and adherence/nonadherence. A 95% credible interval for each estimate is simply the 2.5th and 97.5th percentiles of the 90,000 samples.

As our goal was to determine which factors were associated with adherence/nonadherence, we examined the point estimates and 95% credible intervals for the means, given in Table 3. These estimates from the posterior distribution can be interpreted similarly to the point estimates and confidence intervals for the naive approach shown in Table 2, where an interval containing 0.5 indicates ambiguity or the possibility of no association. The factors in Table 3 have been divided into three groups based on the credible intervals: factors associated with adherence, factors associated with nonadherence, and factors associated with neither adherence nor nonadherence (or both, yielding an estimate near 0.5). Unlike the naive analysis presented earlier, this analysis is not prone to the bias and/or low power that resulted from the construction of normal confidence intervals from small samples.

The research question (identification of factors associated with adherence/nonadherence) is best answered by looking at the factors listed in Table 3. Based on synthesis of information from the 12 reports, we identified four factors that promote adherence, four factors that are unrelated or ambiguously related to adherence, and four factors that hinder adherence.

If we wish to delve more deeply into the relative strength of the relationships, the 90,000 samples can be used to compare the strength of association between the 12 factors and adherence/nonadherence. We can calculate the exact posterior probability that any factor is more closely related to adherence/nonadherence than any other factor by comparing the entire set of samples for the two factors. For example, we might want to examine whether

the literature supported a stronger relationship between adherence and the belief that antiretroviral medications might do good than between adherence and a positive social network. According to Table 3, both credible intervals span values > 0.5 , so both of these factors were associated with adherence. The point estimate for the belief that antiretroviral medications might do good is higher (.96 vs. .85), but we want to have some idea of the chance that there actually is a difference. We calculated the posterior probability that the mean for belief that medications might do good was higher than that for positive social network. To do this, we used the 90,000 samples from the posterior distribution of the population mean. We counted how often (i.e., in how many samples) the sampled value for the belief that antiretroviral medications might do good was higher than that for positive social network.

As it turned out, the sampled value was higher for the belief that antiretroviral medications might do good than for positive social network in 88,811 of the 90,000 samples, or 98.7% of the time. Accordingly, we can be 98.7% confident that the mean for belief that medications might do good is higher than for positive social network. That is, we are 98.7% confident that the evidence is more supportive of a relationship between belief that medications might do good and adherence than between a positive social network and adherence.

This is a type of Bayesian hypothesis testing. It can be thought of more naturally as a hypothesis test if we calculate posterior probability that the mean for belief that medications might do good is less than or equal to that for positive social network, which is simply $100\% - 98.7\% = 1.3\%$. So, there is only a 1.3% chance that the relationship between adherence and positive social network was equal to or stronger than that between adherence and belief in effectiveness. This posterior probability is the Bayesian analog to a p -value, so it is some-times called a *Bayesian p -value* (Gelman et al. 2004). It is the posterior probability that there is no real difference in the population and is often interpreted like the usual (frequentist) p -value by comparing it to a significance level of 5%. If we were doing a Bayesian significance test by that strict criterion, we would say there is evidence to show that the relationship between adherence and belief in effectiveness is stronger than that between adherence and positive social network. However, just as the 5% cutoff for p -values is rather arbitrary for the frequentist approach, it is for the posterior probability as well. A researcher would not be unreasonable to use a higher or lower cutoff, or to simply interpret the probabilities directly with no stated significance level.

This process of comparing samples can be used to assess the relative strength between adherence and any 2 of the 12 factors, but is not entirely necessary. In general, if the credible intervals do not overlap, we can be at least 95% confident that there is a difference between the two means (i.e. the Bayesian p -value will be less than 5%). For example, the credible intervals for belief that medications might do good (.90, 1.0) and having or living with children (.46, .64) do not overlap. Therefore, the posterior probability that the mean for belief that medications might do good is less than or equal to that for having or living with children is no more than 5% (and likely much smaller than 5%). When an exact posterior probability is desired, or when the credible intervals do overlap, a pairwise comparison of the two factors as described above can be performed.

Since non-overlapping credible intervals will always yield a Bayesian p -value of less than 5%, we chose not to calculate Bayesian p -values for these comparisons. Pairwise comparisons were performed for all pairs of variables with overlapping credible intervals. There were 21 of those comparisons. Table 4 contains the results for the 11 comparisons with Bayesian p -values less than .20. The others are omitted from the table for the sake of brevity. Interpreting the Bayesian hypothesis tests with a significance level of 5%, we conclude that belief that medications might do good was significantly more associated with

adherence than any other factors. Positive social network and routinization were not significantly different in their associations. Positive social network was significantly different from having or living with children. Routinization was significantly different from having or living with children and healthy or low virologic load. The two factors associated with neither adherence nor nonadherence (having or living with children and healthy or low virologic load) were not significantly different from each other. None of the 5 factors associated with nonadherence were significantly different from one another.

The Bayesian data augmentation procedure involved estimating a covariance between findings in a study. Converting these findings to correlations, we can get estimates of (and credible intervals for) the relationships between findings within a study. We chose not to discuss the correlations in depth in this paper, but provide a single example here for the interested reader. The data provide evidence that the correlation between the relationship between side effects and adherence and the relationship between a complex regimen and adherence is .25, with a 95% credible interval of (.07, .42). That is, reports citing a relationship between side effects and adherence would tend to report a similar relationship between complex regimen and adherence. Other point estimates of correlations ranged from $-.40$ to $.44$. Future work may further examine the implications of this correlation structure on this or other datasets.

4 Discussion

Those arguing for the applicability of Bayesian analysis to research synthesis often focus solely on the Bayesian ability to incorporate prior beliefs into an analysis- that is, first stating prior beliefs about the study question, then updating those beliefs based on evidence from several research reports. Although the incorporation of prior information is a useful feature of Bayesian analysis, that was not its main attraction here. In this analysis, we used fairly non-informative (objective) priors, and exploited instead the innate ability of the Bayesian approach to borrow information to allow the observed data to inform about missing values.

We have described a Bayesian method that can be applied to the synthesis of results from qualitative and quantitative reports. Unlike Roberts et al. (2002), we incorporated the qualitative and quantitative evidence in the same step of the analysis (rather than using qualitative reports only as prior information). This means that the data from each report, whether qualitative or quantitative, is treated the exact same way in the model. In contrast to Voils et al. (2009), which synthesized data at the participant level by linking findings to the numbers of participants expressing them, we synthesized data at the study level, by designating themes as present or absent, rather than at the subject level whereby findings are linked to numbers of subjects expressing them.

To our knowledge, the current analysis is the first Bayesian synthesis of qualitative and quantitative research that allows for an explicit comparison of the strength of association among a set of factors being studied. We were able to examine the individual and relative effects of 12 different factors on adherence. Roberts et al. (2002) examined multiple factors, but looked at the relationships of each with the outcome as mutually exclusive events (in Bayesian terms, a major limitation of their analysis). Voils et al. (2009) estimated the probability of an association between medication regimen complexity and non-adherence in a single subject, but has not yet been expanded to account for multiple factors simultaneously.

The rigorous analysis did not yield more interesting results than the naive analysis simply by virtue of being Bayesian. The naive analysis was a frequentist (non-Bayesian) analysis, but a

Bayesian analysis with a non-informative prior that did not borrow information across studies would have yielded identical results to the naive frequentist analysis. The frequentist approach was presented because it was simpler in implementation and explanation.

The data matrix was constructed by characterizing the presence (or absence) of themes in each report. Tabulating the results in reports of qualitative and quantitative studies according to the presence or absence of results is not new. For example, Mills et al. (2006b) reviewed the literature with the intent to describe typical barriers to participation in cancer clinical trials. They amassed a list of themes across the reports they reviewed and then categorized each theme as either present or absent in each research report. In contrast to this dichotomous 0/1 scheme, our coding scheme has a distinct advantage in that it distinguishes between variables not studied and variables that were studied, but for which no association was found. In Mills et al. 2006b, unstudied relationships were treated as equivalent to findings of no relationships. This could lead to bias toward relationships that have been more frequently studied. Our use of missing data methods to account for unstudied/unreported factors parallels what is often done in primary studies to account for unmeasured variables.

Although our coding scheme successfully separated findings of no relationship from unstudied findings, it is still limited in that it did not separate factors unassociated with adherence from those factors associated with both adherence and nonadherence. For example, a finding from the qualitative studies was that having or living with children promoted adherence when it gave the woman a purpose in life, but hindered adherence when child care interfered with her ability to care for herself (e.g., Edwards 2006; Powell-Cope et al. 2003). Although it did not happen in this set of reports, if both of those findings were reported in a single study, the relationship would be coded as .5, the same as no association. From a clinical standpoint, this number does not adequately describe the nature of the relationship between having or living with children and adherence.

Tabulation of results in this way lends itself well to the qualitative reports as there is no need to infer information (i.e., actual frequency counts) that is not explicitly present in them (and may not be available from authors). Moreover, this approach is more in line with the qualitative research imperative to include all findings, even those expressed by only one participant. This approach may be less advantageous for quantitative reports, however, as it involves dichotomizing a finding that is generally reported with a p -value, effect size, or some other measure of the strength of effect; such dichotomization results in the loss of information. Future projects may relax the restriction of values to 0, 0.5, and 1, allowing for a more continuous value to capture varying degrees of association.

The data augmentation model we used is certainly not the only model, and different models or different adaptations of this model should be considered in the future. We used a truncated normal likelihood, but we also considered defining a latent variable folded-normal likelihood, where all the mass above one is put on 1 and all the mass below 0 is put on 0, as in Albert and Chib (1993). We chose not to do this because the inflation on 1 and 0 might bias toward strong associations, but in some settings those inflations might be desirable. A potential limitation to flexibility of the data augmentation model is the choice of the inverse-Wishart distribution to model the covariance. In any study where the goal is to draw inferences about specific elements of the covariance matrix, a more flexible distribution should be considered.

Finally, the application of this method did not require both qualitative and quantitative studies, as every study was treated the same. Although we did not include any factors not studied in both quantitative and qualitative reports, this was not a necessary modeling

decision. In fact, the imputation of missing values that borrows information across studies makes this method uniquely suited to the case where a variable is more heavily covered in qualitative or quantitative studies. Future applications of this method need not focus on the qualitative/quantitative nature of each study, but rather seek to include all studies covering variables of interest regardless of the study methodology.

Acknowledgments

This article was supported by a National Institute of Nursing Research, National Institutes of Health (5R01NR004907, June 3, 2005-March 31, 2010) grant (“Integrating qualitative & quantitative research findings”), and with resources and facilities at the Veterans Affairs Medical Center in Durham, NC. Views expressed in this article are those of the authors and do not necessarily represent the Department of Veterans Affairs.

References

- Abel E, Painter L. Factors that influence adherence to HIV medications: Perceptions of women and health care providers. *J. Assoc. Nurses AIDS Care*. 2003; 14:61–69. (*indicates HIV adherence report in Table 1). *. [PubMed: 12953613]
- Albert JH, Chib S. Bayesian analysis of binary and polychotomous response data. *J. Am. Stat. Assoc.* 1993; 88:669–679.
- Barbour RS, Barbour M. Evaluating and synthesizing qualitative research: the need to develop a distinctive approach. *J. Eval. Clin. Pract.* 2003; 9:179–186. [PubMed: 12787181]
- Berry DA. Bayesian clinical trials. *Nat. Rev. Drug Discov.* 2006; 5:27–36. [PubMed: 16485344]
- Casella G, George EI. Explaining the Gibbs sampler. *Am. Stat.* 1992; 46:167–174.
- Chang Y, Voils CI, Sandelowski M, Hasselblad V, Crandell JL. Transforming verbal counts in reports of qualitative descriptive studies into numbers. *West. J. Nurs. Res.* 2009; 31:837–852. [PubMed: 19448052]
- Choi H, Shen R, Chinnaiyan AM, Ghosh D. A latent variable approach for meta-analysis of gene expression data from multiple microarray experiments. *BMC Bioinformatics*. 2007; 8:364. [PubMed: 17900369]
- Dixon-Woods M, Agarwal S, Jones D, Young B, Sutton A. Synthesizing qualitative and quantitative evidence: a review of possible methods. *J. Health Serv. Res. Policy*. 2005; 10:45–53. [PubMed: 15667704]
- Douglass JL, Sowell RL, Phillips KD. Using Peplau’s theory to examine the psychosocial factors associated with HIV-infected women’s difficulty in taking their medications. *J. Theory Constr. Test.* 2003; 7:10–17. *.
- Durante AJ, Bova CA, Fennie KP, Danvers KA, Holness DR, Burgess JD, et al. Home-based study of anti-HIV drug regimen adherence among HIV-infected women: feasibility and preliminary results. *AIDS Care*. 2003; 15:103–115. *. [PubMed: 12655838]
- Eakin JM, Mykhalovskiy E. Reframing the evaluation of qualitative health research: reflections on a review of appraisal guidelines in the health sciences. *J. Eval. Clin. Pract.* 2003; 9:187–194. [PubMed: 12787182]
- Edwards LV. Perceived social support and HIV/AIDS medication adherence among AfricanAmerican women. *Qual. Health Res.* 2006; 16:679–691. *. [PubMed: 16611972]
- Gant LM, Welch LA. Voices less heard: HIV-positive African American women, medication adherence, sexual abuse, and self-care. *J. HIV/AIDS Soc. Serv.* 2004; 3:67–91. *.
- Gelman, A.; Carlin, JB.; Stern, HS.; Rubin, DB. *Bayesian Data Analysis*. 2nd edn. Chapman and Hall/CRC; Boca Raton, Florida: 2004.
- Gerring, J. *Case study research: Principles and practices*. Cambridge University Press; New York: 2007.
- Goldsmith MR, Bankhead CR, Austoker J. Synthesizing quantitative and qualitative research in evidence-based patient information. *J. Epidemiol. Community Health*. 2007; 61:262–270. [PubMed: 17325406]

- Hammersley, M. Systematic or unsystematic, is that the question? Some reflections on the science, art, and politics of reviewing research evidence; Paper presented to Public Health Evidence Steering Group of the Health Development Agency; National Institute for Clinical Excellence; 2002. Available via http://www.nice.org.uk/niceMedia/pdf/sys_unsys_phesg_hammersley.pdf. Cited 2 February, 2009
- Harden A, Thomas J. Methodological issues in combining diverse study types in systematic reviews. *Int. J. Soc. Res. Methodol.* 2005; 8:257–271.
- Horrace WC. Some results on the multivariate truncated normal distribution. *J. Multivar. Anal.* 2005; 94:209–221.
- Howard AA, Arnsten JH, Lo Y, Vlahov D, Rich JD, Schuman P, et al. A prospective study of adherence and viral load in a large multi-center cohort of HIV-infected women. *AIDS.* 2002; 16:2175–2182. *. [PubMed: 12409739]
- Ickovics JR, Wilson TE, Royce RA, Minkoff HL, Fernandez MI, Fox-Tierney R, et al. Prenatal and postpartum zidovudine adherence among pregnant women with HIV results of a MEMS substudy from the Perinatal Guidelines Evaluation Project. *JAIDS: JAIDS.* 2002; 30:311–315. *.
- Jones, DR.; Dixon-Woods, M.; Abrams, K.; Fitzpatrick, R. Meta-analysis of qualitative and quantitative evidence: full report of research activities and results (ESRC ALCD 2 programme). 2005. Available by searching the ESRC social sciences repository. <http://www.esrcsocietytoday.ac.uk>. Cited 6 January, 2009
- Kalichman SC, Rompa D, DiFonzo K, Simpson D, Austin J, Luke W, et al. HIV treatment adherence in women living with HIV/AIDS: Research based on the Information-Motivation-Behavioral Skills Model of Health Behavior. *J. Assoc. Nurses AIDS Care.* 2001; 12:58–67. *. [PubMed: 11486721]
- Little, RJA.; Rubin, DB. Statistical Analysis with Missing Data. 2nd ed. Wiley; Hoboken, NJ: 2002.
- Liu H, Longshore D, Williams JK, Rivkin I, Loeb T, Warda US, et al. Substance abuse and medication adherence among HIV-positive women with histories of child sexual abuse. *AIDS Behav.* 2006; 10:279–286. *. [PubMed: 16501869]
- Mills EJ, Wilson K, Rachlis B, Griffith L, Wu P, Guyatt G, Cooper C. Barriers to participation in HIV drug trials: a systematic review. *Lancet Infect. Dis.* 2006a; 6:32–38. [PubMed: 16377532]
- Mills EJ, Seely D, Rachlis B, Griffith L, Wu P, Wilson K, Ellis P, Wright J. Barriers to participation of clinical trials in cancer: a meta-analysis and systematic review of patient-reported factors. *Lancet Oncol.* 2006b; 7:141–148. [PubMed: 16455478]
- Misener TR, Sowell RL. HIV infected women's decisions to take antiretrovirals. *West. J. Nurs. Res.* 1998; 20:431–447. *. [PubMed: 9686522]
- Mostashari F, Riley E, Selwyn PA, Altice FL. Acceptance and adherence with antiretroviral therapy among HIV-infected women in a correctional facility. *JAIDS: JAIDS.* 1998; 18:341–348. *.
- Murphy DA, Greenwell L, Hoffman D. Factors associated with antiretroviral adherence among HIV-infected women with children. *Women Health.* 2002; 36:97–111. *. [PubMed: 12215006]
- Phillips KD, Moneyham L, Murdaugh C, Boyd MR, Tavakoli A, Jackson K, et al. Sleep disturbance and depression as barriers to adherence. *Clin. Nurs. Res.* 2005; 14:273–293. *. [PubMed: 15995155]
- Pope, C.; Mays, N.; Popay, J. Synthesizing qualitative and quantitative health evidence. Open University Press; Berkshire, England: 2007.
- Powell-Cope GM, White J, Henkelman EJ, Turner BJ. Qualitative and quantitative assessments of HAART adherence of substance-abusing women. *AIDS Care.* 2003; 15:239–249. *. [PubMed: 12856345]
- Richter DL, Sowell RL, Pluto DM. Attitudes toward antiretroviral therapy among African American women. *Am. J. Health Behav.* 2002; 26:25–33. *. [PubMed: 11795602]
- Roberts KJ, Mann T. Barriers to antiretroviral medication adherence in HIV-infected women. *AIDS Care.* 2000; 12:377–386. *. [PubMed: 11091770]
- Roberts K, Dixon-Woods M, Fitzpatrick R, Abrams K, Jones DR. Factors affecting uptake of childhood immunisation: an example of Bayesian synthesis of qualitative and quantitative evidence. *Lancet.* 2002; 360:1596–1599. [PubMed: 12443615]

- Sandelowski M, Voils CI, Barroso J. Comparability work and the management of difference in research synthesis studies. *Soc. Sci. Med.* 2007; 64:236–247. [PubMed: 17029691]
- Sankar A, Luborsky M, Schuman P, Roberts G. Adherence discourse among African-American women taking HAART. *AIDS Care.* 2002; 14:203–218. *. [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 Behav.* 2005; 9:267–279. *. [PubMed: 16088368]
- Siegel K, Gorey E. HIV-infected women: barriers to AZT use. *Soc. Sci. Med.* 1997; 45:15–22. *. [PubMed: 9203266]
- Siegel K, Lekas H-M, Schrimshaw EW, Johnson JK. Factors associated with HIV-infected women's use or intention to use AZT during pregnancy. *AIDS Educ. Prev.* 2001; 13:189–206. *. [PubMed: 11459356]
- Sowell RL, Phillips KD, Seals BF, Misener TR, Rush C. HIV-infected women's experiences and beliefs related to AZT therapy during pregnancy. *AIDS Patient Care STDs.* 2001; 15:201–209. *. [PubMed: 11359662]
- Stone VE, Hogan JW, Schuman P, Rompalo AM, Howard AA, Korkontzelou C, et al. Antiretroviral regimen complexity, self-reported adherence, and HIV patients' understanding of their regimens: survey of women in the HER Study. *JAIDS: JAIDS.* 2001; 28:124–131. *.
- Sutton AJ, Higgins JPT. Recent developments in meta-analysis. *Stat. Med.* 2007; 27:625–650. [PubMed: 17590884]
- Voils CI, Sandelowski M, Barroso J, Hasselblad V. Making sense of qualitative and quantitative findings in mixed research synthesis studies. *Field Methods.* 2008; 20:3–25. [PubMed: 18677415]
- Voils CI, Hasselblad V, Crandell JL, Chang Y, Lee E, Sandelowski M. A Bayesian method for the synthesis of evidence from qualitative and quantitative reports: an example from the literature on antiretroviral medication adherence. *J. Health Serv. Res. Policy.* 2009; 14:226–233. [PubMed: 19770121]
- Wilson TE, Barrón Y, Cohen M, Richardson J, Greenblatt R, Sacks HS, et al. Adherence to antiretroviral therapy and its association with sexual behavior in a national sample of women with human immunodeficiency virus. *Clin. Infect. Dis.* 2002; 34:529–534. *. [PubMed: 11797182]
- Wilson TE, Ickovics JR, Fernandez MI, Koenig LJ, Walter E. Self-reported zidovudine adherence among pregnant women with human immunodeficiency virus infection in four US states. *Am. J. Obstet. Gynecol.* 2001; 184:1235–1240. *. [PubMed: 11349194]
- Wood SA, Tobias C, McCree J. Medication adherence for HIV-positive women caring for children: in their own words. *AIDS Care.* 2004; 16:909–913. *. [PubMed: 15385246]

Table 1
Data matrix relating the factors to nonadherence in the 12 qualitative and 13 quantitative studies

	Side effects/feeling ill	Healthy or low virologic load	More complex regimen	Belief that anti-retroviral medications might do harm	Belief that anti-retroviral medications might do good	Having or living with children	Negative feelings	Disclosure/stigma/threats to confidentiality.	Routinization	Positive social network
Qualitative										
Abel and Painter (2003)	0		0	1	1		0	0	1	
Edwards (2006)								0		1
Gant and Welch (2004)	0		0			1				1
Misener and Sowell (1998)	0			0	1					0.5
Powell-Cope et al. (2003)	0					0	0	0	1	1
Richter et al. (2002)	1				1				1	
Roberts and Mann (2000)	0		0	0	0	0		0		
Sankar et al. (2002)					1			0		
Schrimshaw et al. (2005)	0		0	0	1					
Siegel and Gorey (1997)	0			0						
Siegel et al. (2001)	0	0		0						1
Wood et al. (2004)						1				1
Douglass et al. (2003)	0									1
Durante et al. (2003)								0		1
Howard et al. (2002)	0.5					0.5				
Ickovics et al. (2002)			1							
Kalichman et al. (2001)						1	0		0	
Liu et al. (2006)										0.5
Mostashari et al. (1998)				0.5	1					1
Murphy et al. (2002)					0.5					0.5
Phillips et al. (2005)							0			
Sowell et al. (2001)	0.5									
Stone et al. (2001)							0.5			
Wilson et al. (2002)		1								0.5
Wilson et al. (2001)								1		

Table 2

Naive estimates and confidence intervals relating the factors to adherence/nonadherence

Factor	Number of cases (studies)	Mean (95% confidence interval)	Truncated confidence interval
Belief that antiretroviral medication might do good	7	.93 (.79, 1.07)	(.79, 1)
Positive social network	11	.81 (.67, .97)	(.67, .97)
Routinization	4	.75 (.26, 1.24)	(.26, 1)
Having or living with children	6	.58 (.19, .98)	(.19, .98)
Healthy or low virologic load	2	.50 (−.48, 1.48)	(0, 1)
More complex regimen	5	.20 (−.19, .59)	(0, .59)
Side effects/feeling ill	12	.17 (−.02, .35)	(0, .35)
Disclosure/stigma/threats to confidentiality	7	.14 (−.14, .42)	(0, .42)
Negative feelings	5	.10 (−.10, .30)	(0, .3)
Belief that antiretroviral medications might do harm	6	.08 (−.08, .25)	(0, .25)

Table 3

Point estimates and posterior credible intervals from the Bayesian data augmentation algorithm

	Factor	Posterior mean	95% credible interval	For comparison, the truncated 95% confidence interval from the naive analysis
Factors associated with adherence	Belief that antiretroviral medication might do good	.96	(.90, 1.0)	(.79,1)
	Positive social network	.85	(.76, .94)	(.67,.97)
	Routinization	.80	(.60, .98)	(.26,1)
Factors associated with neither adherence or nonadherence	Having or living with children	.64	(.45, .83)	(.19,.98)
	Healthy or low virologic load	.62	(.29, .90)	(0,1)
Factors associated with nonadherence	More complex regimen	.29	(.11, .48)	(0,.59)
	Disclosure/stigma/threats to confidentiality	.21	(.07, .36)	(0,.42)
	Side effects/ feeling ill	.21	(.09,.32)	(0,.25)
	Negative feelings	.19	(.10,.29)	(0,.3)
	Belief that antiretroviral medication might do harm	.16	(.08, .25)	(0,.25)

Table 4

Selected pairwise comparisons between factors with overlapping credible intervals

Factor 1	Factor 2	Number of samples in which Factor 1 had a higher estimate than Factor 2	Bayesian <i>p</i> -value
Belief that antiretroviral medication might do good	Positive social network	88,811	.01*
Having or living with children	More complex regimen	89,532	.01*
Healthy or low virologic load	Disclosure/stigma/threats to confidentiality	88,773	.01*
Healthy or low virologic load	Side effects/feeling ill	89,103	.01*
Positive social network	Having or living with children	87,388	.03*
Healthy or low virologic load	More complex regimen	86,040	.04*
Positive social network	Healthy or low virologic load	83,360	.07
More complex regimen	Belief that antiretroviral medication might do harm	81,550	.09
Routinization	Having or living with children	77,991	.13
Routinization	Healthy or low virologic load	75,799	.16
More complex regimen	Negative feelings	75,100	.17

Ten pairs of factors with overlapping credible intervals had a Bayesian *p*-value above .20 and are not included in this table Bayesian *p*-value <.05