Developing a Bayesian Statistical Approach to Behavioral Intervention Trials

David Huh

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2012

Reading Committee:

Jane M. Simoni, Chair

Brian P. Flaherty

John Miyamoto

Program Authorized to Offer Degree:

Psychology

University of Washington

Abstract

Developing a Bayesian Statistical Approach to Behavioral Intervention Trials

David Huh

Chair of the Supervisory Committee:
Professor Jane M. Simoni
Department of Psychology

Bayesian statistical methods permit the incorporation of existing knowledge formally into a statistical analysis to generate estimates that leverage both past and present information. Bayesian methods are particularly suitable for intervention research since treatments are generally evaluated over the course of multiple trials. Although Bayesian methods have seen increasing use in the evaluation of biomedical trials, there has been minimal application to date in the behavioral and psychosocial intervention literature. Moreover, Bayesian techniques have been used in an *ad hoc* manner, with some intervention studies leveraging existing data and others utilizing non-informative priors. This study focused on developing a framework for Bayesian analysis that addresses the practical considerations of accumulating data across heterogeneous studies. We used real data from a series of three randomized controlled trials conducted in New York City, Seattle, and Beijing that evaluated behavioral interventions for improving HIV antiretroviral adherence and mental health outcomes. In contrast to previous

literature that has been limited to cross-sectional statistical models, this case study demonstrated the accumulation of data using multilevel regression techniques. In the first set of analyses, we evaluated medication adherence outcomes in the New York City and Seattle studies, the most similar with respect to design and interventions. In the second set of analyses, we evaluated depression outcomes in the full sequence of studies, which introduced the complexity of accommodating a study with substantial differences in methodology. The integration of data from multiple sources led to refined estimates of intervention effectiveness. However, differences arose in the estimates of intervention effectiveness across studies, raising substantive questions about when the decision to aggregate data may be appropriate. We discuss the role of contextual, implementation, and secular effects that may influence an aggregated analysis. This case study illustrated the substantive considerations necessary to support the decision to combine data across studies, and the need for careful review of findings to confirm the appropriateness of pooling using a Bayesian approach.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Number

LIST OF TABLES

Table Number

ACKNOWLEDGEMENTS

The research data for which this dissertation is based was supported by National Institute of Mental Health grants R01 MH58986 and in part by MH074364 and MH074364-S1 to Jane M. Simoni, Ph.D.

OVERVIEW

The Bayesian and classical approaches are the two major branches of statistical inference currently guiding the analysis of randomized controlled trials (RCTs). Although Bayesian inference dates backs back more than two centuries, the practical application of Bayesian statistical models has only become possible in the past three decades, due to advances in computing and mathematical algorithms capable of estimating the models. In the first of a two-part dissertation, I undertake a qualitative review of Bayesian statistical methods in the behavioral and psychosocial intervention literature. The first section of the review provides an overview of the frequentist statistical methods typically used in RCT research. In the second section, the theory behind the frequentist and Bayesian approaches to inference is reviewed. The rigidity of frequentist analyses and their challenging interpretation will be contrasted with the relative flexibility of Bayesian analyses and their more straightforward interpretation. In the third section, I discuss the implications of these two forms of inference with respect to the analysis of randomized controlled trials. In the final section, I review how Bayesian methods have been applied in behavioral and psychosocial intervention research, including their justification and patterns in their application. In the second half of this dissertation, I review the methodological challenges of reconciling heterogeneous studies in a simultaneous analysis using a case study of three real datasets to illustrate a systematic approach to Bayesian analysis using multilevel modeling.

A QUALITATIVE REVIEW OF BAYESIAN STATISTICAL ANALYSIS IN BEHAVIORAL INTERVENTION RESEARCH

Bayesian statistical methods permit the incorporation of existing knowledge formally into a statistical analysis to generate estimates that leverage both past and present information. In recent years Bayesian methods have seen increasing application in the biomedical trial literature (Hobbs & Carlin, 2008) with multiple guidance documents released regarding the implementation of Bayesian analysis in drug and medical device trials (U.S. Food and Drug Administration, 2010a, 2010b). Within psychology, there has been growing interest by cognitive scientists in Bayesian models of perception, learning, and reasoning (Kruschke, 2010), and statistical methodology (e.g., Rouder & Lu, 2005; Rouder, Speckman, Sun, Morey, & Iverson, 2009; Rouder, Sun, Speckman, Lu, & Zhou, 2003; Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010). The strength of the Bayesian analytic process is the accumulation of data in a coherent manner, from the initial piloting of an intervention to subsequent larger scale experiments. The systematic updating of analyses using data from earlier studies has been illustrated in other fields (e.g., Sullivan & Mieczkowski, 2008), however, their adoption in psychosocial and behavioral intervention research has proceeded slowly.

Classical statistics, also known as *frequentist* statistics, have been the standard statistical approach for evaluating randomized trials, (Goodman, 1999). Bayesian methods are more flexible because they approach statistical uncertainty in a more principled manner than frequentist methods (Berry, 1993). A key consequence is that there is no penalty for repeated analysis since Bayesian models are designed to be updated as new data emerges. In the sections that follow I review the basic theory underlying frequentist and Bayesian inference and the implications for intervention research. I then undertake a qualitative review of how Bayesian methods have been used to date in the psychosocial and behavioral intervention literature.

Frequentist versus Bayesian Inference

Frequentist inference. The most common approach to statistical inference in randomized trials is the frequentist perspective, the theoretical basis of classical statistics (Neyman, 1937). The term "frequentist" is connected with the concept of repeated sampling and link between statistical inference and the behavior of samples. Inference in classical statistics assumes that a study could be replicated an infinite number of times, each with a random sample of individuals from the population. In a classical analysis, the mean of the repeated samples all things being equal, reflects a best estimate of the population. The variation around the mean reflects our uncertainty due to sampling. In the frequentist perspective, an objective truth is believed to exist; the aim of the analysis is to reconstruct that truth from a sample of data (Zampetakis, Moustakis, Dewett, & Zampetakis, 2008).

For example, the analysis of a cognitive-behavioral therapy intervention might reveal an average reduction in depression score of 10-points over the control group. That average is an estimate of the true effect of the intervention, representing what happened in a single study. However, the estimate of the intervention effect might be different if the study were repeated again. For instance, the average intervention effect might be 9 points in a second sample, 11 points in a third sample, and 10 points in a forth sample. That is, with each repetition of the intervention, the observed benefit of intervention over control may vary. In the frequentist perspective, variability arises from the process of drawing differing samples, each with limited data.

The cornerstone of frequentist inference, the Central Limit Theorem, states that as successive samples are accumulated, the collection of statistics will approximate a normal distribution, specifically the sampling distribution of the statistic (Howell, 2010). Moreover, the mean of the sampling distribution (i.e., the average across samples) converges to the presumed truth as more samples are collected. The *asymptotic* assumption that a sample increasingly resembles the population with greater sample size is a key aspect of the theory underlying classical statistics (Agresti, 2002). That theory justifies the exploitation of samples to represent the behavior of the population. In general, frequentist statistical methods are highly dependent on asymptotic theory, and in studies with small samples, results may be unreliable unless specialized procedures are utilized. For example, Fisher's Exact test is recommended to evaluate contingency tables with sparse cells (Fisher, 1922) due to a violation of asymptotic assumptions.

Bayesian Inference. The term "Bayesian" is a reference to the Reverend Thomas Bayes who developed Bayes' theorem, a method for calculating the probability of a conclusion given a set of information (Barnard & Bayes, 1958). In Bayesian inference, statistical uncertainty is defined as the relative probability that each possible outcome may be true. Statistical inference is achieved by updating of the probability of each possible outcome in light of relevant new information (Edwards, Lindman, & Savage, 1963). Bayes' theorem provides a mathematical framework for the statistical analysis, such that the "outcome" is any unknown that one wishes to ascertain, such as the strength of an intervention. The "set of information" is then the data collected through scientific study, such as an RCT. Consequently, Bayes' theorem relates a specific hypothesis (e.g., the intervention decreases alcohol consumption by one drink per week) to an explicit probability, from 0 to 100%, that the hypothesis is true. This probability is known as the posterior probability and is the product of three components: (1) the prior, (2) the likelihood, and (3) the normalizing constant. The first two components are the most critical, representing (1) the probability that the hypothesis is true given what was known before the study began (2) the probability that the hypothesis would have yielded the data that were observed. These are known, respectively, as the prior probability of the hypothesis and the likelihood of the data. The third component is a "bookkeeping" adjustment so that the formula yields a percentage that falls between 0 and 100.

Bayes' formula can be easily understood using a diagnostic test analogy (Rosen, 1998). Suppose a certain disorder occurs in 0.5% of the population. A medical test is available that can detect its presence 99% of the time in afflicted patients, but will give a false positive in 5% of administrations. Assuming that the test comes back positive, the actual probability that the patient has the disorder is not 99%, the sensitivity of the test, because the disorder is known to be extremely rare in the population. Factoring in the low base rate of the disorder, the actual probability that the positive test is correct is 91%. Consequently, Bayesian inference can be seen as a means of incorporating base rate information into a statistical analysis.

From the Bayesian perspective, all hypotheses are inherently uncertain and the objective is to summarize the probabilities of the respective hypotheses. Taking the previous intervention example, a Bayesian analysis may reveal a 10-point reduction in depression score over a control group 50% of the time, a 9-point reduction 30% of the time, an 8-point reduction 10%, and so forth. All of these outcomes are assumed to be possible and the ultimate effectiveness of the intervention is thought to be function of chance. However, the intervention is expected to reduce depression scores by 8 to 9 points 80% of the time. The Bayesian approach treats uncertainty as inherent in nature while the frequentist approach assumes that uncertainty arises from the process of observation and collecting samples.

Inference in Randomized Controlled Trials

Frequentist inference. The frequentist approach is exemplified by the use of null hypothesis testing and the interpretation of confidence intervals and *p* values to judge statistical significance (Berry, 1993). In a simple two-arm RCT with a pre- and a post-intervention assessment, a standard null hypothesis would be that the groups performed equally. If the intervention reaches the conventional $p < .05$ threshold, the null hypothesis is rejected and it is concluded that there is a statistically significant difference. A frequent interpretation of $p = .05$ is that there was a 5% chance or less of no treatment effect (Wijeysundera, Austin, Hux, Beattie, & Laupacis, 2009). However, this straight forward interpretation is incorrect. What $p = .05$ actually means is that *assuming* that the intervention was no different than the control, there was a 5% or smaller chance of observing results *as extreme or more extreme* than what was actually observed (Kline, 2004).

The limitations of *p* values have led to calls for emphasizing confidence intervals, since they communicate the magnitude and precision of an effect (Gardner & Altman, 1986). However, clinicians frequently misinterpret confidence intervals to mean that there is 95% probability that the true effect lies within the interval. In reality, the true interpretation is less direct: if the same study was replicated repeatedly and a confidence interval was calculated for each replication, 95% of those intervals would include the “true” intervention effect.

Another implication of the frequentist approach is that inference is tied to the sample size planned in the design of the study. If frequentist theory is followed strictly, then any deviation from the original power calculations can invalidate statistical inference (Berry, 1993, 2005). This arises because the interpretation of the *p* values and confidence intervals are directly tied to the planned sample. As a result, over-enrollment in an RCT increases the probability of Type I error and under-enrollment increases the likelihood of Type II error. Despite these limitations, strengths of frequentist analysis include how it does not require any pre-existing knowledge about the effectiveness of an intervention and that with modern statistical software frequentist statistical models can be run quickly.

Bayesian inference. In a formal Bayesian approach, previous evidence about the effectiveness of an intervention (e.g., pilot data) is first translated into a *prior distribution*, which summarizes the relative probability of each possible intervention effect based on previous knowledge (Sullivan & Mieczkowski, 2008). In the second step, data from the current study are incorporated into a *likelihood distribution*, which summarizes the probability of the data associated with each possible intervention effect. The data from the current study factor into the analyses through this likelihood distribution (Gelman, 2004). In the third step, past and present data are united when the prior and likelihood distributions are combined to form the posterior distribution. The process of "updating" the prior distribution using current data reflects the increase in knowledge that occurs in Bayesian analysis. The posterior distribution summarizes the present state of knowledge, informed by both the most recent work and knowledge from the past.

Bayesian statistical methods that leverage prior data possess greater statistical power to detect intervention effects. Wijeysundera and colleagues (2009) identified 88 parallel-arm studies with dichotomous or time-to-event outcomes and re-analyzed them using Bayesian analysis. In all, 39 studies had positive intervention findings and 49 had negative findings under the original classical analyses. For comparison, the Bayesian re-analyses were conducted separately with non-informative, skeptical, and enthusiastic prior data. Under non-informative priors, the Bayesian results agreed with classical findings in all the positive studies. However, under skeptical and enthusiastic priors, results were different. With a skeptical prior, 13% of positive studies were less effective according to a Bayesian analysis than the original analysis suggested. With an enthusiastic prior, 6% of positive trials were less effective than the original analyses. Alternately, under non-informative and enthusiastic priors, trials that yielded negative findings in the original analyses were shown to be more effective in the Bayesian analyses. With a non-informative prior, 10% of studies with negative findings yielded positive findings under the Bayesian analysis. With an enthusiastic prior, 31% of studies with negative findings were positive under Bayesian analysis.

The emergence of additional findings under a non-informative prior suggests that Bayesian analysis may be more sensitive to intervention effects that are not detectable under classical methods. A Bayesian analysis with a non-informative prior parallels a classical analysis except that a complete probability model of the data is estimated. Given identical data, a full probability model is theoretically more accurate than a partial probability model, suggesting that the Bayesian results are more valid. It must also be noted that a subset of intervention effects deemed positive under a classical analysis were attenuated under a Bayesian analysis with enthusiastic prior information. This suggests that leveraging informative prior information will not always generate positive findings.

Criticisms against the Bayesian approach. Despite the theoretical appeal of Bayesian inference as a means of leveraging historical information to more powerfully evaluate interventions, frequentist approaches predominant due to fundamental issues in the application of Bayesian methods. The use of prior knowledge has been a contentious aspect of Bayesian inference because the choice of data that go into the prior is frequently subjective (Howard, Coffey, & Cutter, 2005). Consequently, the final result of a Bayesian analyses may depend on the choice of which prior data to include or exclude. From the standpoint of statistical power, Bayesian analysis is more advantageous with increasing amounts of previous information. However, leveraging previous information involves the risk of carrying forward the biases of previous studies. Those biases may be due to methodological problems from earlier studies (e.g., poor treatment implementation) or a skew towards evidence in support due due to the “file drawer effect.”

This concern can be partially addressed by conducting sensitivity analyses with multiple priors, an approach long advocated in the biostatistics literature (Spiegelhalter, Freedman, & Parmar, 1994). Analyses can be first performed with a *non-informative prior*, where the results are determined almost exclusively by the study data. Such an analysis closely approximates a traditional frequentist analysis (Bernardo, 1979). Second, a *clinical prior* can be constructed from an earlier study of the intervention (Spiegelhalter et al., 1994), results from meta-analysis (Brophy & Joseph, 1995; Spiegelhalter, Abrams, & Myles, 2004), or more subjectively through a survey of expert opinion (Chaloner & Rhame, 2001). Third, an *enthusiastic prior* could be constructed that reflects a high degree of confidence in the intervention working and minimal likelihood of no benefit or an iatrogenic effect (Diamond & Kaul, 2004; Spiegelhalter et al., 1994). Finally, a *skeptical prior* could be constructed that reflects a low confidence in the intervention’s efficacy, placing a higher burden of evidence on proving the intervention effective.

In summary, Bayesian inference addresses many of the limitations of frequentist methods. First, Bayesian probabilities have a direct interpretation which can be understood by non-statisticians (Berry, 2005). Second, Bayesian analysis allows the incorporation of additional data as they become available without a statistical penalty for repeated analysis. Third, Bayesian methods can be easily used to evaluate the clinical, and not just statistical, significance of interventions. Finally, Bayesian analysis is already used in clinical decision-making to evaluate the cost-effectiveness of interventions (e.g., Johnson-Masotti, Laud, Hoffmann, Hayat, & Pinkerton, 2004; Lambert, Billingham, Cooper, Sutton, & Abrams, 2008; e.g., Martikainen, Valtonen, & Pirttilä, 2004). A use of Bayesian methods to evaluate the interventions themselves would contribute to a consistent framework for clinical decision making.

Bayesian Intervention Analysis in Behavioral and Psychosocial Research

To identify intervention studies that used Bayesian analysis, we conducted a search of the published literature in the Medline and PsychInfo databases. We used the keyword "Bayesian" in combination with the terms: *intervention*, *treatment*, *randomized, controlled trial*, *clinical trial*, or *RCT*. The search was limited to English language articles with no restrictions imposed on the publication date. Studies included in this review evaluated interventions that were (1) behavioral or psychosocial in nature, (2) designed to modify behavioral or psychological outcomes, and (3) targeting a patient or community population. These search criteria permitted a broad array of study designs including randomized trials as well as non-randomized studies, including quasi-experiments and single group designs. The intent was to capture any form of behavioral or psychosocial intervention, to evaluate the use of Bayesian analysis outside of biomedical research. These broad criteria yielded seven articles, suggesting only limited application of Bayesian statistical methods in psychosocial treatment research. However, this small collection of studies illustrates attempts to address common statistical issues encountered by intervention researchers.

Bayesian methodology first appeared in the 1970s for the purpose of evaluating compensatory education programs. An early study by Wang, Novick, Isaacs, and Ozenne (1977) analyzed pilot data from an education program targeted at underperforming students. The design of the study provides insight into why a Bayesian approach was chosen over classical methods. A total of 17 school districts were sampled and a pair of schools was selected within each district according to similar characteristics. These schools were then randomized so that one school in each district became a test site for the compensatory education intervention while the other served as a matched control. To track student performance, standardized achievement tests were given at both the beginning and end of the school year and the outcome was the change in score for each student.

An analysis of covariance model was chosen to assess the average change in performance for intervention versus control participants. Although there was sufficient data for an overall evaluation of intervention effectiveness, the investigators were also interested in a finer grade assessment of how the intervention fared in each of the 34 schools involved in the study. However, with fewer than 20 study participants in a fifth of the schools, a traditional ANCOVA would not have been statistically powered to evaluate differences in intervention effectiveness (Peat & Barton, 2005). Consequently, a Bayesian approach was chosen since a small amount of prior data permitted the intervention analysis to accommodate schools with sparse data. The uninformative prior information incorporated into the analysis did not give preference to either a positive or negative finding. At the substantive level, this allowed a comparison of student outcomes receiving the intervention from different teachers.

A later study conducted by Rubin (1981a) used Bayesian methods to re-assess the effectiveness of a coaching program for improving Scholastic Aptitude Test (SAT) scores previously tested using traditional methods. A classical analysis of variance had been initially used to estimate the effect of SAT coaching in eight separate high schools (Alderman & Powers, 1980). Similar to the Wang and colleagues study (1977), Rubin (1981a) sought to make multiple comparisons across several sites where the intervention had been implemented, but with a similar complication of small samples in specific schools. Rather than using an uninformative prior to facilitate the analysis, he utilized an empirical Bayes approach in which the prior data were derived from the data itself. The results indicated that the coaching program had significantly increased SAT scores in each of the eight schools. Since the traditional approach had answered the question of whether the intervention was effective, Rubin's aim was to illustrate how a Bayesian approach could provide additional possibilities for summarizing intervention performance that would not be possible under classical analysis. In his reanalysis, Rubin demonstrated that by employing Bayesian analysis of variance, it was possible to not only to test the difference between intervention and control, but also to graphically depict the entire distribution of intervention effects, along with the probability of each. Although classical statistical models can be coerced into providing this level of detail is possible with simulation methods (e.g., King, Tomz, & Wittenberg, 2000), this is a natural extension of the information provided by the Bayesian analysis.

This early work demonstrated the utility of Bayesian statistical models for enhancing statistical reporting and tackling challenges such as low sample size. However, in the past three decades only a handful of RCT reports in the psychosocial and behavioral intervention literature have used Bayesian analysis. Within organizational and industrial psychology, Svantec, O'Connell, and Baumgardner (1992) illustrated the flexibility of Bayesian methods for evaluating organizational development (OD) interventions. Hypothetical employee satisfaction data from three time points were created to demonstrate a how a Bayesian *t* test could be used to conduct a simple repeated-measures analysis. This analysis tested the change between time 2 and

3, using the change between time 1 and 2 as the prior data. This formulation was artificial since the data from time 1, 2 and 3 were collected as part of the same hypothetical study. Therefore, none of the observed data were truly “prior” in the sense that it was known before the “study” was conducted. Despite the limitation of their working example, the authors highlighted several advantages of Bayesian methods that applied not only to the evaluation of OD interventions but to other areas of research. First, small samples sizes are common in preliminary studies, but can lead to misleading results if analyzed with classical methods. Second, they noted that differentiating between statistical and clinical significance is a natural extension of Bayesian methods, but that the two are frequently confused when reporting the resulting null hypothesis tests from classical analyses.

Recent Bayesian RCT reports have utilized increasingly sophisticated statistical models. Gill, Reifsnider, and Lucke (2007) used a combination of Bayesian logistic regression and survival analysis to evaluate the effectiveness of a combination prenatal education and postpartum telephone support intervention on the initiation and duration of breastfeeding, respectively. A total of 158 Hispanic women were recruited in a quasi-experimental sampling of two matched maternity clinics, participants from one site receiving the intervention, and the other control. In the first logistic regression analysis, the intervention was associated with two-fold greater odds of breastfeeding initiation. A *post hoc* survival analysis of initiators showed half the quit rate of breastfeeding compared with control participants. Non-informative prior distributions were used for all the models. The authors noted increasing application of Bayesian analysis in nursing research as a broad rationale for its use in their report.

A recent pilot study by Zampetakis, Moustakis, Dewett, and Zampetakis (2008) used Bayesian latent growth modeling (LGM) to evaluate the effectiveness of a one-semester creativity course on self-reported creative behavior. Similar to HLM, the LGM method examines individual change over time, but using a structural equation modeling (SEM) approach (Bollen & Curran, 2006). A total of 94 college engineering students were randomized to a 10-session creativity course that met once weekly or an alternative management course. Creative behavior was assessed via self-report at baseline, 5-week mid-intervention, and 10-week post intervention. As with previous studies, the investigators used sample size to justify their use of a Bayesian approach, noting that any fewer than 200 cases could be considered a small sample in the context of SEM (Weston & Gore, 2006).

Since the intervention was previously untested with no data to indicate its possible effectiveness, Zampetakis and colleagues chose a non-informative prior in their Bayesian model. This prior contributed minimal information to the analyses beyond the data provided by the 94 participants. The results suggested a trend towards more improvement in self-reported creativity behavior among students randomized to the intervention group. However, a limitation of their analysis was that the difference between the intervention and control arm participants was not tested directly. Rather than including intervention group as a predictor, creativity behavior trajectories were examined separately for intervention and control and then later compared. Despite this methodological concern, the Zampetakis et al. study demonstrated the strengths of a Bayesian analysis of pilot data. Rather than focusing on whether the intervention group would improve over time at a $p < .05$ level of significance, the investigators were able to describe the actual probability (62%) that those receiving the creativity intervention improved over time. Alternately, they were able to report the probability that creativity would decrease as a function of the creativity intervention (13%).

Another recent study by Swanson and colleagues (2009) used Bayesian HLM to evaluate the effectiveness of three couples-focused interventions versus no treatment for managing depression and grief in the year following a miscarriage. In contrast to earlier RCT reports that turned to Bayesian analyses due to small sample concerns (Svyantek et al., 1992; Wang et al., 1977; Zampetakis et al., 2008), the study of Swanson and colleagues was statistically powered to conduct either a classical or Bayesian analysis. A total of 340 couples were randomized to nurse counseling, video therapy, or a combination of both. The intervention was administered three times over the course of 11 weeks and depression and grief were assessed at baseline, 3, 5, and 13 months. The results indicated that counseling had the broadest beneficial effect across grief and depression. Although more methodologically vigorous than previous studies, key details regarding the analyses were omitted, specifically the type of prior data that was used. A specific benefit of Bayesian estimation was the ability to describe the probability that each intervention performed better than the alternatives.

The use of Bayesian analysis to impute missing outcome data was illustrated by Joffe and colleagues (2009) in their evaluation of two classroom-based smoking cessation programs among adolescent smokers: Not on Tobacco (NOT) and Kickin' Butts (KB). Eight schools were randomized to one of the group interventions, with a total of 407 high school students participating. Quit status was assessed by self-report and biochemical verification at baseline, post-intervention, and 1, 3, 6, and 12-month follow ups. Notably, the primary analysis was a classically estimated longitudinal regression assessing the relative risk of quitting vs. not quitting at each time point. Bayesian analysis using a non-informative prior was conducted as a secondary approach where all available data, including intervention status and additional smoking-related covariates, were used to impute missing outcome data. The classical analysis without imputation identified higher rates of quitting for one of the interventions (NOT) intervention at 1 month, but no other assessment points. In contrast the Bayesian analysis identified higher quit rates for the NOT program at three out of five time points (post-intervention, 1, and 12 months). Although a direct comparison between the sensitivity of the classical and Bayesian analysis is not possible as they used different sets of covariates, the Joffe and colleagues paper illustrates the advantage of Bayesian estimation in incorporating imputation as a simple extension of a model.

Discussion

Although the application of Bayesian analysis in the behavioral intervention literature has been limited, a number of patterns have emerged in its justification and use within the field. First, the most frequent justification has been the superiority of Bayesian over classical estimation when either the overall sample or individual group sizes were small. Second, the sophistication of statistical models has increased, from cross-sectional techniques in the earliest reports to more sophisticated longitudinal models in the last decade. A common element of both earlier and more recent reports has been a move beyond simple null hypothesis tests to summaries intervention effects in a more intuitive manner. Many focus on the actual probability that the intervention performed better than control, rather than a *p* value.

However, there have been methodological and reporting problems in the literature. A key inconsistency is the selection of prior data, with many studies leveraging historical information (e.g., Sullivan & Mieczkowski, 2008; Svyantek et al., 1992), but others choosing non-informative priors (Wang et al., 1977; Zampetakis et al., 2008), or failing to describe the prior at all (Swanson et al., 2009). None of the reviewed studies conducted sensitivity analyses to compare the effect of different types of prior data on the final results, an approach used in the medical literature (Hobbs & Carlin, 2008).

RCT studies in the behavioral sciences have generally utilized Bayesian techniques in an *ad hoc* manner (e.g., Swanson et al., 2009; Zampetakis et al., 2008). In particular, the choice of prior data, one of the defining elements of the Bayesian approach, has varied greatly among studies. Some studies have utilized only non-informative (e.g., Zampetakis et al., 2008) or only informative (e.g., Svyantek et al., 1992) priors. Moreover, in instances where prior data have been used, few details have been provided to justify their use. In summary, RCT reports in the behavioral intervention literature have only begun to exploit the full potential of Bayesian methods, either to incorporate accumulated knowledge or to test fine-grained hypotheses of clinical effectiveness.

A BAYESIAN APPROACH TO THE LONGITUDINAL ANALYSIS OF RANDOMIZED CONTROLLED TRIALS

The application of Bayesian modeling to evaluate intervention studies is timely given growing interest in the area of integrative data analysis (IDA). Curran and Hussong (2009) define IDA as the simultaneous analysis of two or more samples that have been combined into a single data set. In contrast to meta-analysis, which focuses on synthesizing summary statistics, IDA is premised on utilizing the original datasets. The rationale is to provide a more powerful test of scientific hypotheses that leverages a cumulative base of information. In this respect, the rationale of IDA closely parallels that of Bayesian analysis.

Although IDA techniques have found use in behavioral intervention research, notably the analysis of multisite randomized studies (Raudenbush & Liu, 2000), the analyses are almost universally performed using classical statistical inference. The disadvantage of classical methods is that they incur an increased risk of chance findings (i.e., Type I error) with each repeated analysis. Bayesian inference is the optimal means of undertaking IDA since it is premised upon continual re-analysis as additional data become available.

Integrating Data from Heterogeneous Studies

Despite the theoretical appeal of Bayesian analysis, there are major methodological challenges in aggregating multiple intervention studies. By reconciling methodological differences in a principled manner, the eventual conclusions of a Bayesian analysis are more likely to be interpretable. The objective in this case study is to orient substantive researchers to two broad sources of between study differences: intervention heterogeneity and study design heterogeneity.

Intervention heterogeneity. When consolidating data from multiple RCTs, a theoretical and practical decision must be made as to which intervention groups can be meaningfully combined. Similarly, the equivalence of control comparison conditions requires evaluation. In a prototypical scenario, a newly developed intervention is pilot tested, optimized according to feedback and results, and evaluated in a more extensive investigation if the initial findings show promise. As replications proceed, the intervention may be modified for each study. Small refinements may not alter the essential nature of the intervention, but substantial changes to content, duration, intensity, or delivery method may give rise to a fundamentally different intervention. Additionally different forms of control comparison (e.g., waitlist versus attention) may be utilized across studies and another decision must be made as to whether they can be aggregated.

Where tractable, intervention or control group heterogeneity may be reconciled by identifying a broader definition that subsumes the variations across studies. For example, a set of psychotherapy studies targeting medication adherence in HIV-positive patients may utilize different treatment protocols (e.g., motivational interviewing, skills training), but all could be defined more generally as cognitive behavioral therapy. Additionally, the comparison conditions may all be characterized as “standard of care,” recognizing the presence of geographic differences in usual treatment. Once all the unique intervention and control conditions are identified, a coding scheme must be established that organizes similar and dissimilar interventions for the statistical analysis.

The categorical coding of intervention arms may vary in the original studies. In a two-arm study, intervention A may be denoted by a single indicator variable with the lowest category corresponding with the control condition. A follow-up 2 x 2 factorial trial testing intervention A and B may be represented by separate main effect parameter for each intervention and an interaction term to account for any additional effect when given in combination.

Study design heterogeneity. Differences in sampling criteria, randomization procedures, assessment frequency, and measures are only a few of the factors that require careful consideration when undertaking a single common analysis (Curran & Hussong, 2009; Hofer & Piccinin, 2009). The chosen statistical model must accommodate structural differences between studies. For example, the same intervention may have been evaluated in a single-arm design, a randomized two-arm design with a control comparison, and a quasi-experimental design in which participants chose whether they received the intervention. Assuming random sampling, the intervention groups from the single-arm and two-arm designs could potentially be combined. However, equating RCT participants who were randomized to receive the intervention versus quasi-experimental participants who elected it would require more careful justification.

The shape of longitudinal effects (e.g., planned contrasts, growth curves) must be compatible with both the timing of assessments shape and the shape of trajectories in the respective studies. When assessments are conducted at the same points, longitudinal effects may be testable using categorical contrasts (e.g., baseline vs. follow-up); however, if assessment schedules varied substantially, an HLM approach using growth curves may be the only option.. With two assessment points, only a linear effect is possible, but with three or more data points, the model may need to accommodate curvilinear trends.

Vertical versus Horizontal Data Pooling

The Bayesian literature to date has used a “vertical pooling” approach whereby the effectiveness of an intervention in an initial study is used directly as the prior for analysis of the next investigation (e.g., Abrams, Ashby, & Errington, 1994; Svyantek et al., 1992). To accommodate situations where the past study is less relevant to the current, the variance of the prior data may be inflated by an arbitrary factor so that it has less influence on the final results (Spiegelhalter et al., 1994). The limitation of vertical pooling is that it ignores statistical dependencies that occur when participant outcomes *within* a study are correlated (Rohde, 2008). Greater correlation of participant data from the same study may arise because of environmental similarities that elicit comparable intervention response (Kenny, Mannetti, Pierro, Livi, & Kashy, 2002) or any other factors that vary systematically between studies.

In RCT studies, confounding variables are distributed randomly between intervention arms. However, when multiple RCT data sets are combined, there may be systematic differences between groups randomized in different studies. In theory, one could control directly for the variables associated with differential intervention response. However, study membership serves as a convenient proxy to account for statistical dependencies that arise when combining datasets and is consistent with the nested structure of the aggregated sample (Raudenbush & Liu, 2000). Correct estimation of the intervention effect is contingent upon accounting for known sources of statistical dependency in the data, including the correlation among repeated observations and participants within the same study. Ignoring these statistical dependencies increases the risk of erroneous findings.

Modeling nested data

There are two ways of conceptualizing multiple datasets that have key implications for constructing the statistical model. In the random-effect approach, the datasets are viewed as a collection of random samples from a larger population. This gives rise to a model in which participants are nested within a dataset (i.e., a study). The average effect of an intervention in a random-effect model represents an average for the population. In the fixed-effect approach, the datasets are viewed as only representative of the studies from which they were taken. Modeling multiple studies as a random effect requires a large number of datasets (> 30; Snijders & Bosker, 1999) in order to accurately estimate variance between studies. We focus on the fixed effect approach, since it is better suited for the more common scenario where a limited number of datasets are available, perhaps as few as two. In the fixed effect approach, study membership is a predictor in the statistical model.

Although there are multiple potential dummy coding systems for identifying study membership, the chosen scheme must accommodate interventions that were tested in some studies, but not others. The most familiar approach is traditional dummy coding, where one group is chosen arbitrarily as the reference category and the others are estimated as deviations from that reference. The disadvantage of traditional dummy coding is that a single study may not be an appropriate reference across all interventions. For example, study 1 may include interventions A and B and study 2 may include interventions B and C. In this situation, Study 1 could serve as a reference category for interventions A and B or Study 2 for interventions B and C, but neither study could function simultaneously as a reference group for all three interventions.

The most flexible strategy to avoid reference categories is the use of partition dummy coding (Yip & Tsang, 2007). In partition dummy coding, intervention effects and other parameters are coded as stand-alone quantities rather than as deviations from a reference category. Although traditional and partition category coding are mathematically identical (Yip & Tsang, 2007), partition coding flexibly accommodates studies that evaluated different sets of interventions. Both strategies are a form of mixture modeling, in which participants are divided into separate classes (i.e., studies), resulting in a collection (i.e., mixture) of study-specific statistical estimates. The statistical estimates can then be evaluated separately or re-combined as a weighted average.

This approach shares the same principles as pattern mixture models for missing data (cf. Niaura et al., 2002). In the pattern mixture model approach, the effect of missing data is incorporated as an explanatory variable in a longitudinal outcome analysis. The main effect of this missing data indicator and interactions with this indicator are then included in the analysis. The same techniques can be applied in a Bayesian context, except that the grouping variable defines study membership rather than missing data pattern.

Estimation of Bayesian Models

Markov Chain Monte Carlo. Historically, a major factor limiting the appeal of Bayesian analysis was the difficulty in estimating the models. Until the 1990s, Bayesian inference was relegated to a thought experiment, reflecting the scientific ideal of using statistical analysis to accumulate data across time. Before the availability of modern computing, Bayesian analyses were limited to models with an exact mathematical solution. This relegated the universe of Bayesian inference to the simplest models whose integral equations could be solved using calculus. This meant that models of any complexity, including the multilevel models increasingly used for RCT data, were still out of reach.

A seminal paper by Gelfand and Smith (1990) demonstrated how the technique of Gibbs sampling, a form of Markov Chain Monte Carlo (MCMC), could be utilized to *simulate* statistical estimates for Bayesian models whose integral equations were too complex to be solved mathematically. The serendipitous timing of this paper with the modern computing revolution meant that, for the first time, a procedure was now available for estimating Bayesian models along with the processing power necessary to exploit that procedure. In the past two decades, Bayesian analysis has gradually become more accessible through not only specialized software such as WinBUGS (Lunn, Thomas, Best, & Spiegelhalter, 2000), but also widely used statistical packages such as SAS and SPSS Amos, which include MCMC estimation capability.

One limitation of MCMC simulation is that the software is still maturing. Estimation algorithms for classical analyses have benefitted from decades of development and refinement. On a mathematical level, the integral calculus equations underlying Bayesian analyses are more inherently complicated than the corresponding differential equation of a classical analysis. Software packages utilizing MCMC approaches are continually improving, but remain less robust than classical techniques. Current MCMC software packages require careful configuration and close monitoring to ensure that the estimations converge properly (Gelman & Hill, 2006).

Expectation-Maximization Algorithm. The MCMC approach to Bayesian estimation involves directly simulating the entire distribution of statistical parameters in a model. After generating a substantial number of simulated parameters, summary statistics such as means and standard deviations can be applied to describe the posterior distribution. In contrast, the expectation-maximization approach (EM) involves an iterative procedure that traces the distribution of the statistical model until a posterior mode, is identified. Key advantages of the EM algorithm include its computational efficiency and stability compared with MCMC simulation. A disadvantage of the EM approach is that that if the posterior distribution is multi-modal, the algorithm may converge to a local mode, but not the global mode. Additionally, the standard EM algorithm provides an estimate of the posterior mode without elucidating the variance around that estimate.

However, EM estimation can be extended to derive the entire posterior distribution. Although modifications to the EM algorithm have been proposed (Wei & Tanner, 1990), a practical option is a Bayesian bootstrap (Rubin, 1981b) in which a large number of simulated datasets are generated using resampling. The Bayesian model can then be applied to each resampled dataset using standard EM to derive simulated estimates of the posterior mode. The distribution of statistical estimates across the bootstrapped datasets provides an estimate of the posterior distribution. An additional advantage of pairing EM estimation with bootstrapping is that the standard errors are resistant to violations of parametric statistical assumptions, such as normal distribution of the outcome, or equal variance of the outcome across groups.

Data Augmentation Priors. An alternative to MCMC estimation is to leverage classical analytic routines to estimate Bayesian models. An approximate Bayesian analysis involves incorporating the prior information as additional observations of data and estimating the "augmented" dataset as one would the corresponding classical statistical model. The use of *data augmentation priors* permits Bayesian analyses to be carried out in any statistical package (Greenland & Christensen, 2001). Recall that a Bayesian statistical model is a combination of past knowledge, represented by a prior distribution, and the most recent data, represented by a likelihood distribution. In conjugate Bayesian analyses, the prior and likelihood belong to the same family of probability distributions. In this situation, the prior and observed data are interchangeable, meaning that the sequence in which information is combined mathematically makes no difference in the ultimate result. Consequently, an analysis in which past knowledge is explicitly placed into the prior distribution is equivalent to a reorganized analysis in which that same prior information is collated with the observed data. The data augmentation approach exploits the mathematical interchangeability of the prior and likelihood distributions. Although data augmentation is considered an "approximate" Bayesian technique, in practice the difference with the results from MCMC and other "true" Bayesian techniques are trivial, making them appropriate for applied research. Importantly, data augmentation has become an acceptable approach to Bayesian analysis (Greenland & Christensen, 2001).

Prior Distributions

The prior distribution, or probability, has both a philosophical and practical interpretation. Philosophically, the prior distribution establishes a base rate for each possible value of a statistical parameter. An uninformative prior assigns an equal or near-equal probability for any value of the parameter. On a practical level, an equivalent interpretation of the prior distribution is that it is a sample of data. Consequently, the prior and likelihood are equivalent to two samples of data that are pooled in a Bayesian analysis.

Informative priors can be constructed from information from (1) past research, (2) expert opinion (Chaloner & Rhame, 2001), or (3) subjective assumptions (Mayo & Gajewski, 2004). Non-informative priors are the most commonly used in intervention analyses, which lead to results driven almost entirely by the observed data, with little to no influence from past knowledge. Although the updating of statistical inference is the philosophical objective of Bayesian analysis, there are few examples in the intervention literature where actual data is leveraged. More commonly, Bayesian methodology is the basis of interim analyses *within* the same intervention trial as new data arrives (Spiegelhalter et al., 1994).

Intervention studies leveraging past data have done so with relatively simple cross-sectional models. For example, a common scenario is to assess the odds ratio of successful outcomes in the treatment versus control conditions (e.g., Sullivan & Mieczkowski, 2008). The odds ratio from the first stage analysis is then utilized as the prior for the second stage analysis, and so forth. The advantage of this scenario is that the model is simple enough so that the posterior distribution after each stage of analysis can be expressed in a closed-form integral.

Because the posterior distribution can be integrated using calculus, the results of each analysis can be easily converted into the prior for the next stage analysis

There is no literature to date elucidating how information can be transmitted across successive analyses in even a simple multilevel model, such as a repeated measures analysis of variance. The posterior distribution of a Bayesian multilevel model cannot be expressed in a clean mathematical form (i.e., a closed-form integral), so it must be estimated via simulation or data augmentation priors. However, there is no direct way of translating a simulated posterior distribution into an exact mathematical form that can be used as the prior distribution for the next analysis.

Recall that data augmentation priors can be utilized to estimate Bayesian models in non-Bayesian software. However, they also serve as a practical method to avoid error in translating the posterior distribution of an earlier stage analysis to a prior distribution for the subsequent. In the data augmentation approach, observations from the earlier study are bundled with the follow-up study (Gelman & Hill, 2006), and analyzed with a non-informative prior distribution. In conjugate Bayesian analyses, this is equivalent to updating an informative prior distribution based upon the earlier study with the most recently collected.

Aims of the Present Study

The objective of the present study is to demonstrate a Bayesian approach to integrative data analysis that addresses the practical considerations of combining data from heterogeneous studies. In contrast with previous literature which has been limited to cross-sectional statistical models, our investigation demonstrates how data can be accumulated from study to study using multilevel regression techniques. In the context of RCT research, techniques for longitudinal data analysis may be the most appropriate because they permit the estimation of treatment effects (i.e., group differences) across multiple time-points within a single statistical model.

We use actual data from a series of three RCTs evaluating behavioral interventions for improving HIV antiretroviral adherence and mental health outcomes. The first study, Project HAART (Simoni, Pantalone, Plummer, & Huang, 2007), was a two-arm RCT conducted in the Bronx, New York that evaluated a peer support intervention. The second study, Project Promoting Adherence for Life (PAL; Simoni et al., 2009), was a follow-up 2 × 2 factorial RCT conducted in Seattle that evaluated both peer *and* pager support interventions. The third study, Project Adherence Research in China (ARC; Simoni et al., 2011), was a two-arm RCT conducted in Beijing that evaluated an enhanced intervention where a choice was offered between an alarm device, nurse counseling, or both.

We perform two sets of Bayesian multilevel analyses, beginning with a first stage analysis of the New York study and incorporating the Seattle and Beijing data, in second and third stage analyses. In the first set of analyses, we evaluate medication adherence outcomes in the New York and Seattle studies, the most similar with respect to design and interventions. In the second set of analyses, we evaluate depression outcomes in the full sequence of studies. For both sets of analyses, we evaluate each of the individual studies using classical statistical estimation to assess the sensitivity of the Bayesian results. It was hypothesized that the intervention effects would be more precisely estimated (i.e., have smaller variances), in the second and third stage analyses which leveraged data from each of the preceding studies.

Methods

This is a secondary analysis of three NIH-supported RCT studies conducted in New York City (Simoni et al., 2007); Seattle (Simoni et al., 2009); and Beijing (Simoni et al., 2011). Each study evaluated one or dual interventions for improving antiretroviral (ARV) adherence. All of the three studies included a standard of care (SOC) control group. Interventions provided to patients were delivered in addition to their normal medical care. A synopsis of each study follows, including an overview of the study procedures and summary of the interventions.

Bronx, New York City: Project HAART

Procedures. The sample consisted of 136 patients recruited from the adult HIV primary care outpatient clinic at Jacobi Medical Center, a public institution serving mainly indigent, ethnic minority individuals. Eligible individuals were currently on an ARV regimen, proficient in English, and without dementia or psychosis. Participants ranged in age from 19 to 60 years (M = 43) at baseline and 55% male. The sample included 46% African Americans, 44% Hispanics, 7% Caucasians, and 3% who specified "other" or mixed racial heritages. With respect to education, 44% had some high school or less, 48% had a high school degree or GED, and 8% had a college or associates degree. Data were collected between March 2000 and September 2002.

Participants were randomized to either peer support (n = 71) or standard of care (n = 65) with follow-up interviews occurring at baseline, 3, and 6 months.

Intervention. The three-month peer support intervention consisted of two parts: six twice-monthly one-hour group meetings at the clinic of all peers and actively enrolled participants (i.e., "peer meetings") in addition to weekly phone calls from peers to participants who were assigned to them individually by study staff. HIV-positive clinic patients who were actively engaged in their treatment served as "peers" who provided medication-related social support in the peer support condition. Peers were selected on the basis of their regular attendance of clinic appointments, consistent maintenance of high adherence, and strong social skills. These peers received an initial training and committed to ongoing supervision.

Seattle, Washington: Project PAL

Procedures. The sample consisted of 224 patients recruited from the adult HIV primary care outpatient clinic at Harborview Medical Center, a public institution serving mainly indigent, ethnic minority individuals in Seattle, Washington. Eligible individuals were initiating or switching at least two ARV medications, living within a pager service area, proficient in English, and without dementia or psychosis. Participants ranged in age from 19 to 60 years (M = 40) at baseline and were mostly (76%) male. The sample included 30% African Americans, 11% Hispanics, 47% Caucasians, and 12% who specified *other* or *mixed* racial heritages. With respect to education, 21% had some high school or less, 70% had a high school degree or GED and 9% had a college or associates degree. Data collection occurred between March 2003 and May 2007.

Participants were randomized to one of four study arms: peer support (n = 57), pager support (n = 56), both peer and pager (n = 54), or standard of care (n = 57). Data were collected at baseline (2 weeks for adherence), 3, 6, and 9 months.

Interventions. The three-month peer support intervention was identical in format to the version given in the Bronx, New York and included six twice-monthly one-hour peer meetings and to weekly phone calls placed by peers to participants. The only refinement involved training the peers in a more structured assessment of potential barriers to adherence (e.g., lack of medication knowledge, poor relationship with health providers).

The three month pager intervention consisted of electronic reminders sent out when it was time to take every dose of antiretroviral medications. During the intervention phase, the participants were asked to wear the pager at all times when awake as much as feasible. A confirmation response was requested to confirm receipt and acknowledgement of each page. Three types of supplementary messages were also sent including (1) ARV education, (2) jokes or thoughts of the day; and (3) adherence assessments. Page frequency was gradually reduced in the final month of intervention to minimize a rebound in non-adherence.

Beijing, China: Project ARC

Procedures. The sample consisted of 70 adult patients recruited from Beijing's Ditan Hospital, the premier treatment center for infectious diseases in China. Eligible participants were initiating ARV treatment, proficient in Mandarin, had a CD4 count lower than 350 cells/mm^3, were willing to and physically capable of attending follow-up visits at the hospital, and were not cognitively-impaired or psychotic. Participants ranged in age from 21 to 55 years (M = 36) at baseline, were primarily Han Chinese (96%), and mostly male (81%). Eligible individuals were initiating or switching at least two ARV medications, proficient in English, and without dementia or psychosis. Participants were generally well educated, with 40% having a high school degree and 9%, a college degree; 47% were currently unemployed. Data collection occurred between December 2006 and March 2008.

Participants were randomized to either enhanced intervention (n = 36) or standard of care (n = 34) with were interviewed at baseline, 3, and 6 months.

Interventions. Participants assigned to the enhanced intervention chose between a reminder alarm alone (n = 8), nurse counseling alone (n = 10), or both a reminder alarm and nurse counseling (n = 18). Enhanced intervention participants who selected the alarm device were first offered the option of their cell phone as the electronic reminder. If the participant did not have a cell phone or preferred not to use it, they were provided with a small battery-powered reminder device. The three-month nurse counseling intervention consisted of three one-hour sessions delivered at baseline, 5 and 9 weeks. The intervention was delivered by a bachelor's level nurse and utilized a manualized protocol. A cognitive-behavioral, problem-solving approach was used and included motivational, educational, and skill-building components specific to ARV treatment.

Measures

Medication adherence. Past three-day dose adherence was assessed using a modified version of the Adult AIDS Clinical Trial Group Instrument (Chesney et al., 2000). For each medication, separate questions assessed the number of doses missed yesterday, the day before yesterday, and three days ago. Prescribed regimen was ascertained by referencing the three days prior to the assessment date with information from patient medical charts. From these data, we determined the three-day dose adherence across all ARV medications, calculated as the fraction of doses taken (doses prescribed - doses missed) over the total number of doses prescribed during the period. Past three-day dose adherence data were available in the New York and Seattle studies.

Depression. Participants completed the Centers for Epidemiological Studies Depression Scale, a non-diagnostic screening measure for assessing depressive symptomatology in the past week. The New York and Seattle studies utilized the full 20-item instrument (CESD-20; Radloff, 1977) whereas the Beijing study utilized a Mandarin-language version of the abridged 10-item instrument (CESD-10; Andresen, Malmgren, Carter, & Patrick, 1994). In both versions, participants indicate the frequency that specific depressive symptoms (e.g., "I was bothered by things that usually don't bother me") were experienced, from 0 ["*rarely or none of the time* (*less than one day in the past week*)"] to 3 ["most or all of the time (5-7 days in the past week)"]. Total scores on the CESD-20 ranged from 0 to 60, with higher scores reflecting greater depressive symptomatology. CESD-10 scores were rescaled to match the range of the full instrument. The CESD-20 scale has demonstrated validity, internal consistency, and test-retest reliability (8-week interval, $r = .59$) (Radloff, 1977). The Chinese CESD-10 scale has demonstrated high predictive validity compared with the full CESD ($\kappa = .75 – .84$), high internal consistency ($\alpha = 0.78$), and stable test-retest reliability (3-year interval, $r = 0.44$) (Boey, 1999).

Statistical Analysis

All statistical analyses were performed in the **R** programming environment (R Development Core Team, 2008) and the WinBUGS statistical software (Lunn et al., 2000). Exploratory analyses were first conducted to assess the distribution, inter-correlation, and longitudinal trajectories of the medication adherence and depression outcomes to inform specification of the statistical models.

A sequence of Bayesian models was used to integrate data from the (1) New York, (2) Seattle, and (3) Beijing intervention studies. Separate sets of analyses were conducted for the (a) medication adherence and (b) depression outcomes. The medication adherence analyses were limited to the New York and Seattle data, where patients were randomized to receive a specific single or combination of interventions (*peer support* and/or *reminders*). The depression analysis integrated data from all three sites, including Beijing were participants were randomized to receive a choice of interventions (*reminders* and/or *nurse counseling*).

Intervention coding. The three distinguishable interventions evaluated were *peer support* (New York, Seattle), *electronic reminders* (Seattle, Beijing), and *nurse counseling* (Beijing). In the Beijing study, the distinction between receiving electronic reminders and/or nurse counseling constituted a *post hoc* contrast and not a true randomized comparison. The delivery mechanism of the electronic reminders varied by study; in Seattle, a pager was provided whereas in Beijing, alarms were programmed into a participant's cell phone or an alarm device.

Intervention assignment was represented using traditional dummy coding (Cohen, 2003) with standard of care defined as the reference category. Additionally, two dual intervention combinations were administered, including peer support with reminders and reminders with nurse counseling. To represent the dual intervention groups, we also calculated the peer support × reminder and reminder × nurse counseling interactions. In this coding scheme, the “main effect” term of each intervention represented the effect of the intervention in isolation and the interaction term represented any decrement or improvement in outcomes when two interventions were provided in tandem.

The medication adherence analysis included only the peer support, reminder, and peer support × reminder terms, whereas the depression analysis also included the nurse counseling and reminder × nurse counseling terms.

Multilevel generalized linear model. Multilevel generalized linear modeling was used to assess the effect of intervention on medication adherence, as a percentage of doses taken. We conducted the multilevel analysis using the robust logistic regression approach of Gelman, Jakulin, Pittau, and Su (2008). Robust regression models have been recommended for fractional data in order to derive regression estimates that are resistant to the heteroskedasticity inherent in the data (Papke & Wooldridge, 2008). For routine application, Gelman et al. (2008) recommended a *t* distribution with a scale of 1 and degrees of freedom of 2.5, which is equivalent to assuming one-half addition success and one-half additional failure in a logistic regression. To achieve robust inference, the priors for all regression coefficients were defined as *t* distributions centered at zero with scale and degrees of freedom parameters of 10, a less informative distribution and consequently a more conservative choice than Gelman et al.’s default recommendation.We used the EM algorithm to derive point estimates (i.e., posterior mode) of the regression coefficients, estimated using the `bayesglm` function from the `arm` package in `R` (Gelman et al., 2010).

Bayesian bootstrapping was used to simulate the full posterior distribution (Rubin, 1981b) of the regression coefficients. Bayesian bootstrapping involves three steps: (a) generating a random probability weight for each observation (a) randomly sampling with replacement from the original data with the random weights, (b) calculating the statistic of interest using the generated data, and (c) repeating the process to a desired degree of precision. This iterative procedure approximates the posterior distribution of a desired statistic (e.g., regression coefficient). We performed 20,000 bootstrap simulations using the `boot` package in `R`. We chose this approach rather than MCMC as we found MCMC with WinBUGS converged slowly to the posterior distribution when estimating logistic regressions of fractional outcomes. Specifically, the MCMC algorithms necessitated substantial tuning and adjustment to achieve, whereas the bootstrapping-based approach yielded convergence across models with minimal to no adjustment.

The regression model combining the New York and Seattle studies (Ferron, 1997) is detailed in Table 1. To accommodate non-linear adherence trajectories identified in exploratory analysis, the longitudinal time effect was categorized into three contrasts of baseline versus 3, 6, and 9 months. In the first stage model, adherence in the New York study was regressed on peer support × New York; time (month 3 vs. 0, 6 vs. 0) × New York; peer support × time × New York. In the second stage analysis, adherence in the New York and Seattle studies was regressed on peer support × New York; peer support × Seattle; reminder × Seattle; peer support × reminder × Seattle; time (month 3 vs. 0, 6 vs. 0) × New York; time (month 3 vs. 0, 6 vs. 0) × Seattle; peer support × time (month 3 vs. 0, 6 vs. 0) × New York; peer support × time (month 3 vs. 0, 6 vs. 0, 9 vs. 0) × Seattle; and peer support × reminder × time (month 3 vs. 0, 6 vs. 0, 9 vs. 0) × Seattle. For the second stage model, data from the New York study were incorporated directly into the analysis as data augmentation priors. The statistical tests of the interventions were the peer support × time (month 3 vs. 0, 6 vs. 0, 9 vs. 0); reminder × time (month 3 vs. 0, 6 vs. 0, 9 vs. 0); and peer support × reminder × time (month 3 vs. 0, 6 vs. 0, 9 vs. 0) interactions.

Multilevel linear modeling. Multilevel linear modeling was used to assess the association of intervention with depressive symptomatology. We conducted the multilevel analysis using linear mixed effects regression (LMM), also known as random effects or varying intercept and/or slope models (Gelman & Hill, 2006). The LMM regressions were specified with random intercepts, analogous to repeated-measures analysis of variance. We estimated the model using MCMC simulation with the `R2WinBUGS` package, an `R`-based interface to the WinBUGS statistical program.

The full regression model combining the New York and Seattle studies is detailed in Table 2 (Ferron, 1997). The longitudinal time effect was parameterized as a linear slope. In the first stage model, depressive symptomatology in New York was regressed on peer support × New York, time, peer support × time × New York. In the second stage analysis, depressive symptomatology in New York and Seattle was regressed on peer support × New York, peer support × Seattle, reminder × Seattle, peer support × reminder × Seattle, time × New York, time × Seattle, peer support × time × New York, peer support × time × Seattle, and peer support × reminder × time × Seattle. In the third stage analysis, depressive symptomatology in New York, Seattle, and Beijing was regressed on peer support × New York, peer support × Seattle, reminder × Seattle, peer support × reminder × Seattle, time × New York, time × Seattle, time × Beijing, peer support × time × New York, peer support × time × Seattle, peer support × reminder × time ×

Seattle, reminder × nurse counseling × time × Beijing. For the second and third stage models, data from earlier stage models were incorporated directly into the most current analysis as data augmentation priors. The statistical tests of the interventions were the peer support × time, reminder × time, nurse counseling × time, peer support × reminder × time, and reminder × nurse counseling × time interactions.

Aggregate estimates. Study-specific intervention effects were estimated in each stage of the medication adherence and depression symptom analyses using partition dummy codes (Yip & Tsang, 2007). At each stage of the analyses, we calculated an aggregate effect for each parameter, by averaging across the posterior distributions of the available study-specific effects. Each study specific effect was weighted according to the sample size of the study to create a weighted aggregate effect.

Results

Preliminary analyses

Bivariate correlations, means, standard deviations, skew, and kurtosis statistics by study at each assessment for the medication adherence and depression outcomes are provided in Tables 3 and 4, respectively.

Medication adherence. An examination of the means and standard deviations indicated that average medication adherence was stable from baseline to 3 months in New York (78 to 80%) and Seattle (92 to 89%). Between 3 and 6 months, adherence decreased 8% in both studies from 80 to 72% in New York and from 89% to 81% in Seattle. Adherence remained stable from 6 to 9 months (84%) in Seattle, where 9-month data was available. Individual series plots indicated that raw adherence trajectories varied from linear to curvilinear.

The distribution of medication adherence was multi-modal with the highest concentrations of adherence data at 0% and 100%. Adherence was leptokurtic across time points, with a range of 2 to 4 in New York and 5 to 12 in Seattle, reflecting a high probability of extreme values, a pattern consistent with previous investigations (Pearson, Simoni, Hoff, Kurth, & Martin, 2007). Adherence assessed closer in time (e.g., baseline and 3 months) was more highly correlated than when separated by one or two assessments (e.g., baseline and 9 months), indicating a decay in correlation across greater intervals.

Depressive symptoms. An examination of the means and standard deviations indicated that average depressive symptomatology in the New York and Seattle studies was stable across assessments, varying from 17 to 20 and 23 to 25, respectively. In Beijing, average depressive symptoms decreased between baseline and 3 months from 22 to 16, and 6 months remained stable at 15. Individual series plots indicated that raw depressive symptom trajectories were generally linear in all three studies.

Depressive symptomatology was approximately normal in distribution with skew of < 1 and kurtosis ranging from 2 to 3. Within each study, the correlation of depression data was similar across time points, averaging .69, .55, and .49 in New York, Seattle, and Beijing, respectively.

Medication Adherence in New York and Seattle

Beta coefficients, standard deviations, 95% Bayesian confidence intervals, and odds ratios for the two-stage Bayesian multilevel logistic analysis of medication adherence are reported in Table 5.

Bayesian models. To facilitate comparison, posterior means and Bayesian confidence intervals for the intervention estimates from each stage of analysis, and separately for each study, are plotted in Figure 1.

Peer support. At 3 months compared with baseline, there was equivocal evidence for an effect of peer support on medication adherence in the New York-specific model (*B* = -0.28, *95% CI* = -1.27 – 0.69) with the treatment estimates of highest probability centered near a zero effect, but stronger evidence in the Seattle-specific model (*B* = 1.67, *95% CI* = 0.38 – 2.87) with the treatment estimates of highest probability excluding zero and averaging a five-fold greater posterior odds of 100% adherence. In the second-stage model combining New York and Seattle, there was strong evidence for the effect of peer support with the treatment estimates of highest probability excluding zero and averaging a two-and-a-half-fold greater odds of 100% adherence (*B* = 0.94, *95% CI* = 0.09 – 1.80) at 3 months versus baseline. The posterior distribution of the estimate was more precise in the second-stage model (*SD* = 0.44) than either the New York-specific (*SD* = 0.50) or Seattle-specific (*SD* = 0.63) models.

At 6 months compared with baseline, there was weak evidence for a rebound in medication non-adherence following discontinuation of peer support in both the New York-specific (*B* = -0.69, *95% CI* = -1.67 – 0.26) and Seattle-specific (*B* = -1.10, *95% CI* = -2.84 – 0.47) models with the treatment estimates of highest probability centered at reduced odds of 100% adherence, but including a zero effect. In the second stage model combining New York and Seattle, there was weak evidence for the rebound in medication non-adherence at 6 months versus baseline (*B* = -0.94, *95% CI* = -2.04 – 0.13) with the treatment estimates of highest probability centered at reduced odds of 100% adherence but including a zero effect. The posterior distribution of the estimate in the second stage model (*SD* = 0.54) was comparable in precision to the New York-specific model (*SD* = 0.49), but greater in precision than the Seattle-specific model (*SD* = 0.82).

At 9 months compared with baseline, where only the Seattle study contributed data, the Seattle-specific (*B* = 0.97, *95% CI* = -0.51 – 2.40) and second stage New York and Seattle (*B* = 0.99, *95% CI* = -0.49 – 2.40) models produced consistent posterior estimates. In both models, there was weak evidence for an effect of peer support on medication adherence with the treatment estimates of highest probability centered at improved odds of 100% adherence, but including a zero effect.

Electronic reminders. The Seattle-specific and second stage New York and Seattle models produced consistent posterior estimates for the effect of reminders, where only the Seattle study contributed data. At 3 months compared with baseline, there was weak evidence for an effect of reminders on medication adherence in the Seattle-specific (*B* = 1.23, *95% CI* = -0.05 – 2.47) and second stage New York and Seattle (*B* = 1.24, *95% CI* = -0.01 – 2.50) models with the treatment estimates of highest probability centered at improved odds of 100% adherence but including a zero effect. At 6 months compared with baseline, there was equivocal evidence for an effect of reminders in the Seattle-specific (*B* = -1.01, *95% CI* = -2.79 – 0.62) and second-stage New York and Seattle (*B* = -1.01, *95% CI* = -2.79 – 0.56) models with the treatment estimates of highest probability including a zero effect. At 9 months compared with baseline, there was weak evidence for an effect of reminders in the Seattle-specific (*B* = 1.36, *95% CI* = -0.15 – 2.78) and second stage New York and Seattle (*B* = 1.37, *95% CI* = -0.14 – 2.78) models with the treatment estimates of highest probability centered at improved odds of 100% adherence but including a zero effect.

Interaction of Peer support and Electronic reminders. The Seattle-specific and second stage New York and Seattle models produced consistent posterior estimates for the interactive effect of peer support and reminders, where only the Seattle study contributed data. At 3 months compared with baseline, there was equivocal evidence for an interactive effect of peer support and reminders on medication adherence in the Seattle-specific (*B* = -0.97, *95% CI* = -2.78 – 0.87) and second stage New York and Seattle (*B* = -0.97, *95% CI* = -2.81 – 0.86) models with the interaction estimates of highest probability including a zero effect. At 6 months, there was equivocal evidence for an interactive effect of peer support and reminders on medication adherence in the Seattle-specific (*B* = 1.35, *95% CI* = -0.70 – 3.49) and second stage New York and Seattle (*B* = 1.34, *95% CI* = -0.61 – 3.53) models with the interaction estimates of highest probability including a zero effect. At 9 months, there was equivocal evidence for an interactive effect of peer support and reminders on medication adherence in the Seattle-specific (*B* = -1.29, *95% CI* = -3.33 – 0.84) and second stage New York and Seattle (*B* = -1.32, *95% CI* = -3.33 – 0.81) models with the interaction estimates of highest probability including a zero effect.

Classical models. Classical and Bayesian estimates are plotted in Figure 3. In the study-specific models, classical mean estimates of the intervention effects were close or identical to the Bayesian mean estimates. Bayesian confidence intervals based on a conservative, weakly-informative prior (Gelman et al., 2008) were larger than the corresponding frequentist confidence intervals.

Depression in New York, Seattle, and Beijing

Beta coefficients, standard deviations, and 95% Bayesian confidence intervals for the three-stage Bayesian multilevel linear analysis of depressive symptoms are reported in Table 6.

Bayesian models. To facilitate comparison, posterior means and Bayesian confidence intervals for the intervention estimates, from each stage of analysis and separately for each study, are plotted in Figure 2.

Peer support. There was weak evidence for an iatrogenic effect of peer support on depressive symptom change in the New York-specific model (*B* = 1.64, *95% CI* = -0.05 – 3.40) with the treatment estimates of highest probability centered at slower reduction in depressive symptoms but including a zero effect. In contrast, there was evidence for a beneficial effect of peer support in the Seattle-specific model (*B* = -1.57, *95% CI* = -2.92 – -0.11) with the treatment estimates of highest probability excluding zero and averaging a one-and-a-half point faster reduction every three months in depressive symptoms. In the second stage model combining New York and Seattle, there was equivocal evidence for the effect of peer support on depressive symptom change (*B* = -0.43, *95% CI* = -1.58 – 0.70) with the treatment estimates of highest probability including a zero effect. In the third-stage New York, Seattle, and Beijing model, where no additional peer support data was added, the posterior distribution for the peer support estimate remained constant (*B* = -0.36, *95% CI* = -1.48 – 0.78). The precision of the posterior distribution of the estimate increased from the first-stage (*SD* = 0.88) to second-stage (*SD* = 0.58) model, and remained constant in the third-stage (*SD* = 0.58) model.

Electronic reminders. There was weak evidence for an effect of reminders on depressive symptom change in the Seattle-specific model (*B* = -1.35, *95% CI* = -2.74 – 0.04) with the treatment estimates of highest probability centered at a faster reduction in depressive symptoms, but including a zero effect. There was equivocal evidence in the Beijing-specific model (*B* = 1.54, *95% CI* = -3.06 – 5.79) with the treatment estimates of highest probability including a zero effect. In the second-stage New York and Seattle model, where reminder data was introduced, there was weak evidence for an effect of the intervention on depressive symptom change (*B* = -1.34, *95% CI* = -2.66 – 0.05) with the treatment estimates of highest probability centered at a faster reduction in depressive symptoms, but including a zero effect. In the third-stage New York, Seattle, and Beijing model, evidence for an effect of reminders on depressive symptom change (*B* = -0.65, *95% CI* = -2.07 – 0.80) became more equivocal with the treatment estimates of highest probability converging to a zero effect. The precision of the posterior distribution of the reminder estimate remained constant from the second-stage (*SD* = 0.69) to third-stage (*SD* = 0.74) models. The reminder estimate in the Seattle-specific model (*SD* = 0.71) was more precise than in the Beijing-specific model (*SD* = 2.27), indicating that the third-stage estimate was driven by the Seattle data with minimal influence from the Beijing data.

Nurse counseling. There was equivocal evidence for an effect of the nurse counseling intervention, which was evaluated in the Beijing study, on depressive symptom change in the Beijing-specific model (*B* = -1.12, *95% CI* = -5.18 – 3.08) with the treatment estimates of highest probability including a zero effect. In the third-stage model, where data on nurse counseling was introduced, there was equivocal evidence for an effect of the intervention on depressive symptom change (*B* = -1.10, *95% CI* = -4.97 – 2.99) with the treatment estimates of highest probability including a zero effect. The precision of the posterior distribution of the nurse counseling estimate was consistent between the third-stage (*SD* = 2.03) and Beijing-specific (*SD* = 2.10) models.

Interaction of Peer support and Electronic reminders. There was evidence for an interactive effect of peer support and reminders, which were combined in the Seattle study, on depressive symptom change in the Seattle-specific model (*B* = 2.06, *95% CI* = 0.07 – 4.07) with the interaction estimates of highest probability centered at slower reduction in symptoms. In the second-stage New York and Seattle, where data on peer support and reminders in combination were introduced, there was evidence for an interactive effect on depressive symptom change (*B* = 2.05, *95% CI* = 0.12 – 3.99) with the interaction estimates of highest probability centered at slower reduction in symptoms. In the third-stage New York, Seattle, and Beijing model, where no additional peer support and reminder data was added, the posterior distribution of the interaction estimate remained constant (*B* = 2.05, *95% CI* = 0.13 – 3.91). The precision of the posterior distribution of the interaction estimate remained constant from the second-stage (*SD* = 0.99) to third-stage (*SD* = 0.97) models.

Interaction of Electronic reminders and Nurse counseling. There was equivocal evidence for an interactive effect of reminders and nurse counseling, which were combined in the Beijing study, on depressive symptom change in the Beijing-specific model (*B* = -0.50, *95% CI* = -6.94 – 5.62) with the interaction estimates of highest probability including a zero effect. In the third-stage model, where data on reminders and nurse counseling in combination was introduced, there was equivocal evidence for an interactive effect on depressive symptom change (*B* = -0.52, *95% CI* = -6.57– 5.56) with the interaction estimates of highest probability including a zero effect. The precision of the posterior distribution of the interaction estimate was consistent between the third-stage (*SD* = 3.11) and Beijing-specific (*SD* = 3.21) models.

Classical models. Classical and Bayesian estimates are plotted in Figure 4. Across study-specific models, classical mean estimates of the intervention effects were close or identical to the Bayesian mean estimates. Bayesian confidence intervals based on a non-informative prior were close or identical to the frequentist confidence intervals.

Discussion

This study evaluated a Bayesian statistical approach to intervention analysis in which data from previous investigations are leveraged to inform the analysis of studies to follow. The objective was to illustrate the theoretical considerations and mechanics of combining data from heterogeneous studies, using actual data from three RCTs evaluating behavioral interventions for improving HIV antiretroviral adherence and mental health. The accumulation of knowledge was illustrated using multilevel modeling, increasingly used for the evaluation of longitudinal studies. We improve upon previous approaches, which have assumed the statistically interchangeability of data from different studies are by accounting for site-specific effects in all analyses. The integration of data from multiple sources led to refined estimates of intervention effectiveness. The advantages were most apparent in cases where the range of probable intervention effects included a zero effect when each study was evaluated independently. However, differences arose in the estimates of intervention effectiveness across studies which raise key substantive questions about the reasons behind that variation and when the decision to aggregate data may be appropriate.

In the first set of analyses, we evaluated medication adherence outcomes in the New York and Seattle studies, the most similar with respect to design and interventions. Both studies evaluated a nearly identical version of a peer support intervention for improving HIV antiretroviral compliance. The Seattle study also introduced a reminder intervention that was not tested in the New York Study. We performed two sets of multilevel logistic regressions, beginning with a first stage analysis of the New York study and incorporating the Seattle data in a second stage analysis. The aggregated estimates of the peer support effect at 3, 6 and 9 months, based upon data from both New York and Seattle, were similar or greater in precision than estimates from the individual studies. Specifically, the peer support estimates of highest probability were centered at and included a zero effect on medication adherence at 3 months versus baseline in New York, but were centered at a five-fold greater odds of 100% adherence at 3 months versus baseline, excluding a zero effect, in Seattle. The aggregated peer support estimate was smaller in magnitude than in the Seattle analysis, but had greater precision and a smaller confidence interval. The estimates of the reminder intervention did not change when the two studies were analyzed simultaneously, indicating that data was successfully pooled in one intervention without biasing the evaluation of another.

In the second set of analyses, we evaluated depression outcomes in the full sequence of studies, which introduced the complexity of accommodating a study with substantial differences in study methodology. Whereas the New York and Seattle studies involved a fully randomized design with an identical intervention common to both, participants in the Beijing study were randomized to standard of care or a choice of interventions, including a form of electronic reminder that differed from the version used in the Seattle study. In the initial New York study, peer support participants had marginally poorer improvement in depression over time compared with standard of care, however in the later Seattle study, peer support was associated with a reduction in depressive symptomatology. In the second stage analysis combining both studies, the mean effect of peer support converged to zero. Electronic reminders, which were introduced in the Seattle study, were associated with a marginally higher reduction in depression over time compared, and in the later Beijing study, they were unassociated with depressive symptoms. Similar to the finding with peer support, in the final stage analysis the effect of reminders also converged to zero. An alternating valence of effect can indicate a net effect of zero, but could also arise from differential implementation of the interventions or contextual differences in how different populations responded to the intervention.

It is important to recognize that the interpretation of a Bayesian analysis depends on the comparability of the data sources to be pooled and the extent to which choice to combine is sufficiently justified. Differences in the intervention estimates across studies can arise from a number of sources including, but not limited to, implementation, contextual, and secular effects. Implementation effects result from variation in the intervention due to between study differences in treatment protocol, interventionist skill level, duration and intensity of treatment, intervention fidelity, and variation in study methodology (e.g., outcome measures). Contextual effects arise when the populations of each study respond differently to the intervention. Participant characteristics, such as race or ethnicity, age, culture, and mental and physical health, have the potential of moderating the effectiveness of the intervention. These characteristics may vary by geographical location or be driven by recruitment procedures and eligibility criteria (e.g., a tightly screened efficacy study sampling individuals with less co-occurring pathology compared with an effectiveness study sampling broadly from the community).

Contextual effects arise from an interaction of individual difference and sociological factors, few of which may have been objectively assessed beyond basic demographic characteristics. Although geography can serve as a proxy for context, it may be misleading since distinct subpopulations could be sampled from the same geographic area (e.g., a low-income clinic versus a private hospital). Secular effects are a specific contextual effect connected with response differences due to time. Time can be defined at the individual level, such as time spent with a condition, a newly diagnosed HIV patient versus one who has lived with the diagnosis for many years. It can also be defined by calendar time, such as a sample of newly diagnosed HIV patients in the 1990s versus the 2010s. Consequently, secular effects may be attributable to cohort differences (e.g., earlier diagnosis and improved health among newer patients) that may influence intervention response.

In practice, interventions are rarely replicated under identical conditions meaning contextual and implementation effects are likely the rule rather than the exception. Although this can be viewed as a complicating factor, it is a natural consequence of the intervention development process. However, while some variation between treatment estimates may be unavoidable as interventions and samples change across studies, at what point does pooling become invalid because the treatment itself is no longer the same or because the populations in the respective studies responded to the treatment in distinct ways? The substantive decision to aggregate data across studies must be defended in light of contextual, implementation, and secular effects that may influence the interpretability of the aggregated analysis and subsequent inference.

The present study provided a case exemplar where a compelling *a priori* case could be made for pooling two of the three studies (New York and Seattle) and a weaker case for the third (Beijing). The New York and Seattle studies were the most compatible in that they were both fully randomized designs, shared an intervention that was administered under nearly identical protocols, and used similar outcome measures. The primary implementation differences in the Seattle study were a factorial design involving an additional intervention and a longer follow up period. Thus, the basic structure of the New York study could be considered as nested within the Seattle design. In this context, the New York and Seattle studies could be modeled with parallel regression parameters and the corresponding estimates (i.e., peer support vs. standard of care in New York and Seattle) combined with a minimum of assumptions. Contextually, both the New

York and Seattle studies were conducted at outpatient HIV clinics serving a diverse, indigent patient population. However, many of the first fixed dose combinations, which combine multiple antiretroviral drugs in a single pill, received FDA approval in the early 2000s, so participants in the Seattle study, conducted between 2003 and 2007, benefited from wider availability of less burdensome medication regimens. Thus a secular difference may have made the experiences of receiving HIV treatment less comparable for the cohort of patients in the New York versus Seattle studies.

There were greater substantive differences in the implementation of the Beijing study versus the U.S. studies, including randomization to a choice of interventions, use of an alarm device instead of a pager, and an alternate set of medication adherence and depression measures. Integrating the Beijing study into the analysis of depression outcomes involved subsuming the pagers provided in Seattle and the alarm devices in the Beijing study under a more general class of "electronic reminders," equating participants randomly assigned to an intervention with those who received a choice, and using two variations of the CES-D scale as the outcome. Moreover there were important contextual differences in the Chinese sample including greater HIV stigma. a barrier to medication adherence (Rintamaki, Davis, Skripkauskas, Bennett, & Wolf, 2006), structural differences in medical care delivery, and secular differences in the HIV epidemic in China (Starks et al., 2008) compared with the United States. Because of these factors, the third stage analysis of the depression outcome that incorporates the Beijing data warrants more cautious interpretation.

Collectively, the integration of multiple intervention studies entails a number of methodological considerations. The groundwork involves establishing a case for combining studies, which necessitates a comparison of how the interventions are implemented in each study as well as a review of contextual factors that could affect the response to intervention and establishing an *a priori* rationale for aggregating across the studies. From there, a taxonomy is developed of which interventions can be meaningfully combined across a set of studies. A second aspect involves accommodating study design differences, such as defining longitudinal trajectories in a manner compatible with varying assessment schedules and allowing for differing sets of interventions across studies. A third aspect is controlling for the nesting of data within study, and recombining those stratified estimates. A fourth aspect is accumulating data accurately in a longitudinal Bayesian model, such as through the use of data augmentation priors.

The common approach to accumulating data in a Bayesian framework assumes exchangeability between the prior and likelihood. This is mathematically equivalent to a classical model in which data from two sources are pooled in a single analysis without controlling for nesting or other variables that may systematically differ. Although Bayesian techniques were used to estimate all statistical models, basic principles of model building generally apply regardless of whether the model is approached from a Bayesian versus classical perspective. Consequently, the integration of multiple datasets in a Bayesian analysis does not absolve the researcher from the responsibility of accounting for greater correlation of data from the same sites.

It is important to consider the limitations of the approach that we detail. First, our approach involves controlling for study-specific effects in a Bayesian statistical model and calculating a weighted combination those individual estimates to produce a single index of intervention effectiveness. However, if the statistical estimates are substantially different between sites, a single index may not be appropriate and it may be necessary to examine the study-specific estimates individually. In the same way that the *a priori* decision to combine studies is initially justified on substantive and theoretical grounds, a post hoc decision to reject data pooling would also need to be theory driven. Using our case example, the initial decision to combine peer support data from the New York and Seattle studies was justified on the grounds of the close comparability of the intervention protocols and general comparability of the outpatient contexts in which both studies were conducted, despite being conducted in different cities. The clearly equivocal effect of peer support in the New York study compared with the clearly positive effect in the Seattle study could be used as grounds to reconsider contextual effects as having a stronger role than initially anticipated. For example, we may theorize that peer support may not have been effective for the types of antiretroviral regimens prescribed in the early 2000s. Thus, we may decide that Seattle study may be a more accurate reflection of how peer support performs on more current regimens and not leverage New York data. In this scenario, there is less advantage to Bayesian analysis as information would not be pooled so the precision of the results would be similar or identical to a classical analysis. If compelling arguments can be made both for and against pooling, a reasonable compromise may be to report both the study-specific and aggregated results, with the appropriate qualifiers for interpreting the aggregated results.

A second limitation is the complexity of the software, which has been a major barrier to the adoption of Bayesian analysis. The analyses we demonstrated were performed using freely available packages for the `R` statistical software. With minimal modification, all of the analyses detailed in this report could also be performed in SPSS, SAS, Stata, and other commercial packages by using an approximate Bayesian approach (Greenland & Christensen, 2001) in which informative and non-informative prior data are incorporated into the analysis as additional observations in the dataset. Follow-up research to examine the use of classical statistical software to estimate Bayesian models would be useful since the literature to date has focused on simulation based methods such as MCMC.

Theoretically and practically, Bayesian analysis is an attractive option for evaluating intervention studies. Estimates from the Bayesian models have a direct interpretation as the probability that the interventions are within a range (i.e., 95% confidence interval) of effectiveness, as opposed to the indirect interpretation of a classical model that a replication of the study would give rise to an estimate of the intervention within a range of uncertainty. Analyses can be updated over time, from the initial pilot investigation to later efficacy and effectiveness studies, leveraging all past sources of data without a penalty for repeat use of the same data. The approach detailed in the present study provides a means of incorporating the principles of meta-analysis into all stages of intervention evaluation, with as few as two studies. This case study illustrated the substantive considerations necessary to support the decision to combine data across studies, and careful review of the findings to confirm the appropriateness of pooling. As the field of psychological intervention research moves towards integrative data analysis, Bayesian inference affords an elegant framework for synthesizing multiple sources of information.

REFERENCES

Abrams, K. R., Ashby, D., & Errington, D. (1994). Simple Bayesian analysis in clinical trials: A tutorial. *Controlled Clinical Trials*, *15*, 349–359. doi:10.1016/0197-2456(94)90032-9

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York: Wiley.

Alderman, D. L., & Powers, D. E. (1980). The effects of special preparation on SAT-verbal scores. *American Educational Research Journal*, *17*, 239. doi:10.3102/00028312017002239

Andresen, E. M., Malmgren, J. A., Carter, W. B., & Patrick, D. L. (1994). Screening for depression in well older adults: Evaluation of a short form of the CES-D (Center for Epidemiologic Studies depression scale). *American Journal of Preventive Medicine*, *10*, 77–84.

Barnard, G. A., & Bayes, T. (1958). Studies in the history of probability and statistics: IX. Thomas Bayes's essay towards solving a problem in the doctrine of chances. *Biometrika*, *45*, 293–315. doi:10.2307/2333180

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, *41*, 113–147.

Berry, D. A. (1993). A case for Bayesianism in clinical trials. *Statistics in Medicine*, *12*, 1377–1393. doi:10.1002/sim.4780121504

Berry, D. A. (2005). Clinical trials: Is the Bayesian approach ready for prime time? Yes! *Stroke*, *36*, 1621–1622. doi:10.1161/01.STR.0000170637.02692.14

Boey, K. W. (1999). Cross-validation of a short form of the CES-D in Chinese elderly. *International Journal of Geriatric Psychiatry*, *14*, 608–617. doi:10.1002/(SICI)1099-1166(199908)14:8<608::AID-GPS991>3.0.CO;2-Z

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. Hoboken, NJ: Wiley-Interscience.

Brophy, J. M., & Joseph, L. (1995). Placing trials in context using Bayesian analysis. *JAMA*, *273*, 871–875. doi:10.1001/jama.1995.03520350053028

Chaloner, K., & Rhame, F. S. (2001). Quantifying and documenting prior beliefs in clinical trials. *Statistics in Medicine*, *20*, 581–600. doi:10.1002/sim.694

Chesney, M. A., Ickovics, J. R., Chambers, D. B., Gifford, A. L., Neidig, J., Zwickl, B., & Wu, A. W. (2000). Self-reported adherence to antiretroviral medications among participants in HIV clinical trials: The AACTG adherence instruments. *AIDS Care*, *12*, 255–266. doi:10.1080/09540120050042891

Cohen, J. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Psychology Press.

Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods*, *14*, 81–100. doi:10.1037/a0015914

Diamond, G. A., & Kaul, S. (2004). Prior convictions: Bayesian approaches to the analysis and interpretation of clinical megatrials. *Journal of the American College of Cardiology*, *43*, 1929–1939. doi:10.1016/j.jacc.2004.01.035

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review*, *70*, 193–242. doi:10.1037/h0044139

Ferron, J. (1997). Moving between hierarchical modeling notations. *Journal of Educational and Behavioral Statistics*, *22*, 119–123. doi:10.2307/1165241

Fisher, R. A. (1922). On the interpretation of χ^2 from contingency tables, and the calculation of P. *Journal of the Royal Statistical Society*, *85*, 87–94. doi:10.2307/2340521

Gardner, M. J., & Altman, D. G. (1986). Confidence intervals rather than P values: Estimation rather than hypothesis testing. *British Medical Journal (Clinical Research Ed.)*, *292*, 746–750. doi:10.1136/bmj.292.6522.746

Gelman, A. (2004). *Bayesian data analysis*. Boca Raton: CRC press.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models* (1st ed.). Cambridge University Press.

Gelman, A., Jakulin, A., Pittau, M. G., & Su, Y. S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, *2*, 1360–1383. doi:10.1214/08-AOAS191

Gelman, A., Su, Y. S., Yajima, M., Hill, J., Pittau, M. G., Kerman, J., & Zheng, T. (2010). *arm: Data analysis using regression and multilevel/hierarchical models*.

Gill, S. L., Reifsnider, E., & Lucke, J. F. (2007). Effects of support on the initiation and duration of breastfeeding. *Western Journal of Nursing Research*, *29*, 708–723. doi:10.1177/0193945906297376

Goodman, S. N. (1999). Toward evidence-based medical statistics. 1: The P value fallacy. *Annals of Internal Medicine*, *130*, 995–1004.

Greenland, S., & Christensen, R. (2001). Data augmentation priors for Bayesian and semi-Bayes analyses of conditional-logistic and proportional-hazards regression. *Statistics in Medicine*, *20*, 2421–2428. doi:10.1002/sim.902

Hobbs, B. P., & Carlin, B. P. (2008). Practical Bayesian design and analysis for drug and device clinical trials, *54*, 54–80. doi:10.1080/10543400701668266

Hofer, S. M., & Piccinin, A. M. (2009). Integrative data analysis through coordination of measurement and analysis protocol across independent longitudinal studies. *Psychological methods*, *14*, 150–164. doi:10.1037/a0015566

Howard, G., Coffey, C. S., & Cutter, G. R. (2005). Is Bayesian analysis ready for use in phase III randomized clinical trials? Beware the sound of the sirens. *Stroke*, *36*, 1622–1623. doi:10.1161/01.STR.0000170638.55491.bb

Howell, D. C. (2010). *Statistical methods for psychology* (7th ed.). Belmont: Wadsworth.

Joffe, A., McNeely, C., Colantuoni, E., An, M.-W., Wang, W., & Scharfstein, D. (2009). Evaluation of school-based smoking-cessation interventions for self-described adolescent smokers. *Pediatrics*, *124*, e187–e194. doi:10.1542/peds.2008-2509

Johnson-Masotti, A. P., Laud, P. W., Hoffmann, R. G., Hayat, M. J., & Pinkerton, S. D. (2004). A Bayesian approach to net health benefits: An illustration and application to modeling HIV prevention. *Medical Decision Making*, *24*, 634–653. doi:10.1177/0272989X04271040

Kenny, D. A., Mannetti, L., Pierro, A., Livi, S., & Kashy, D. A. (2002). The statistical analysis of data from small groups. *Journal of Personality and Social Psychology*, *83*, 126–137. doi:10.1037/0022-3514.83.1.126

King, G., Tomz, M., & Wittenberg, J. (2000). Making the most of statistical analyses: Improving interpretation and presentation. *American Journal of Political Science*, *44*, 347–361. doi:10.3886/ICPSR01255.v1

Kline, R. B. (2004). Beyond significance testing: Reforming data analysis methods in the social sciences. *Washington, DC: American Psychological Association*.

Kruschke, J. K. (2010). What to believe: Bayesian methods for data analysis. *Trends in Cognitive Sciences*, *14*, 293–300. doi:10.1016/j.tics.2010.05.001

Lambert, P. C., Billingham, L. J., Cooper, N. J., Sutton, A. J., & Abrams, K. R. (2008). Estimating the cost-effectiveness of an intervention in a clinical trial when partial cost information is available: A Bayesian approach. *Health Economics*, *17*, 67–81. doi:10.1002/hec.1243

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D. J. (2000). WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, *10*, 325–337. doi:10.1023/A:1008929526011

Martikainen, J., Valtonen, H., & Pirttilä, T. (2004). Potential cost-effectiveness of a family-based program in mild Alzheimer's disease patients. *The European Journal of Health Economics*, *5*, 136–142. doi:10.1007/s10198-003-0214-1

Mayo, M. S., & Gajewski, B. J. (2004). Bayesian sample size calculations in phase II clinical trials using informative conjugate priors. *Controlled Clinical Trials*, *25*, 157–167. doi:16/j.cct.2003.11.006

Neyman, J. (1937). Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transactions of the Royal Society of London A*, *236*, 333–380. doi:10.1098/rsta.1937.0005

Niaura, R., Spring, B., Borrelli, B., Hedeker, D., Goldstein, M. G., Keuthen, N., DePue, J., et al. (2002). Multicenter trial of fluoxetine as an adjunct to behavioral smoking cessation treatment. *Journal of Consulting and Clinical Psychology*, *70*, 887–896. doi:10.1037//0022-006X.70.4.887

Papke, L. E., & Wooldridge, J. M. (2008). Panel data methods for fractional response variables with an application to test pass rates. *Journal of Econometrics*, *145*, 121–133. doi:10.1016/j.jeconom.2008.05.009

Pearson, C. R., Simoni, J. M., Hoff, P., Kurth, A., & Martin, D. (2007). Assessing antiretroviral adherence via electronic drug monitoring and self-report: An examination of key methodological issues. *AIDS and Behavior*, *11*, 161–173. doi:10.1007/s10461-006-9133-3

Peat, J. K., & Barton, B. (2005). *Medical statistics: A guide to data analysis and critical appraisal*. Wiley-Blackwell.

R Development Core Team. (2008). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*, 385–401. doi:10.1177/014662167700100306

Raudenbush, S. W., & Liu, X. (2000). Statistical power and optimal design for multisite randomized trials. *Psychological Methods*, *5*, 199–213. doi:10.1037/1082-989X.5.2.199

Rintamaki, L. S., Davis, T. C., Skripkauskas, S., Bennett, C. L., & Wolf, M. S. (2006). Social stigma concerns and HIV medication adherence. *AIDS Patient Care and STDs*, *20*, 359–368. doi:10.1089/apc.2006.20.359

Rohde, P. D. (2008). Statistical analysis of group-administered intervention data: Re-analysis of two randomized trials. *Psychotherapy Research*, *18*, 365–376. doi:10.1080/10503300701796992

Rosen, K. H. (1998). *Discrete Mathematics and its Applications* (4th ed.). McGraw-Hill Science/Engineering/Math.

Rouder, J. N., & Lu, J. (2005). An introduction to Bayesian hierarchical models with an application in the theory of signal detection. *Psychonomic Bulletin & Review*, *12*, 573–604. doi:10.3758/BF03196750

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*, 225–237. doi:10.3758/PBR.16.2.225

Rouder, J. N., Sun, D., Speckman, P. L., Lu, J., & Zhou, D. (2003). A hierarchical bayesian statistical framework for response time distributions. *Psychometrika*, *68*, 589–606. doi:10.1007/BF02295614

Rubin, D. B. (1981a). Estimation in parallel randomized experiments. *Journal of Educational Statistics*, *6*, 377–401. doi:10.2307/1164617

Rubin, D. B. (1981b). The Bayesian bootstrap. *The Annals of Statistics*, *9*, 130–134. doi:10.1214/aos/1176345338

Simoni, J. M., Chen, W.-T., Huh, D., Fredriksen-Goldsen, K. I., Pearson, C. R., Zhao, H., Shiu, C.-S., et al. (2011). A Preliminary randomized controlled trial of a nurse-delivered medication adherence intervention among HIV-positive outpatients initiating antiretroviral therapy in Beijing, China. *AIDS and Behavior*, *15*, 919–929. doi:10.1007/s10461-010-9828-3

Simoni, J. M., Huh, D., Frick, P. A., Pearson, C. R., Andrasik, M. P., Dunbar, P. J., & Hooton, T. M. (2009). Peer support and pager messaging to promote antiretroviral modifying therapy in Seattle: A randomized controlled trial. *Journal of Acquired Immune Deficiency Syndromes*, *52*, 465–473. doi:10.1097/QAI.0b013e3181b9300c

Simoni, J. M., Pantalone, D. W., Plummer, M. D., & Huang, B. (2007). A randomized controlled trial of a peer support intervention targeting antiretroviral medication adherence and depressive symptomatology in HIV-positive men and women. *Health Psychology*, *26*, 488–495. doi:10.1037/0278-6133.26.4.488

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks: Sage.

Spiegelhalter, D. J., Abrams, K. R., & Myles, J. P. (2004). *Bayesian approaches to clinical trials and health-care evaluation*. Chichester: Wiley.

Spiegelhalter, D. J., Freedman, L. S., & Parmar, M. K. B. (1994). Bayesian approaches to randomized trials. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, *157*, 357–416. doi:10.2307/2983527

Starks, H., Simoni, J., Zhao, H., Huang, B., Fredriksen-Goldsen, K., Pearson, C., Chen, W.-T., et al. (2008). Conceptualizing antiretroviral adherence in Beijing, China. *AIDS care*, *20*, 607–614. doi:10.1080/09540120701660379

Sullivan, C. J., & Mieczkowski, T. (2008). Bayesian analysis and the accumulation of evidence in crime and justice intervention studies. *Journal of Experimental Criminology*, *4*, 381–402. doi:10.1007/s11292-008-9062-4

Svyantek, D. J., O'Connell, M. S., & Baumgardner, T. L. (1992). Applications of Bayesian methods to OD evaluation and decision making. *Human Relations*, *45*, 621–636. doi:10.1177/001872679204500605

Swanson, K. M., Chen, H.-T., Graham, J. C., Wojnar, D. M., & Petras, A. (2009). Resolution of depression and grief during the first year after miscarriage: A randomized controlled clinical trial of couples-focused interventions. *Journal of Women's Health*, *18*, 1245–1257. doi:10.1089/jwh.2008.1202

U.S. Food and Drug Administration. (2010a, February). Adaptive design clinical trials for drugs and biologics. Retrieved from http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM201790.pdf

U.S. Food and Drug Administration. (2010b, February 5). Guidance for the use of Bayesian statistics in medical device clinical trials. WebContent. Retrieved March 27, 2012, from http://www.fda.gov/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm071072.htm

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage–Dickey method. *Cognitive Psychology*, *60*, 158–189. doi:10.1016/j.cogpsych.2009.12.001

Wang, M. M., Novick, M. R., Isaacs, G. L., & Ozenne, D. (1977). A Bayesian data analysis system for the evaluation of social programs. *Journal of the American Statistical Association*, *72*, 711–722. doi:10.2307/2286451

Wei, G. C. G., & Tanner, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association*, *85*, 699–704. doi:10.2307/2290005

Weston, R., & Gore, P. A. (2006). A brief guide to structural equation modeling. *The Counseling Psychologist*, *34*, 719–751. doi:10.1177/0011000006286345

Wijeysundera, D. N., Austin, P. C., Hux, J. E., Beattie, W. S., & Laupacis, A. (2009). Bayesian statistical inference enhances the interpretation of contemporary randomized controlled trials. *Journal of Clinical Epidemiology*, *62*, 13–21. doi:10.1016/j.jclinepi.2008.07.006

Yip, P. S. L., & Tsang, E. W. K. (2007). Interpreting dummy variables and their interaction effects in strategy research. *Strategic Organization*, *5*, 13–30. doi:10.1177/1476127006073512

Zampetakis, L. A., Moustakis, V., Dewett, T., & Zampetakis, K. (2008). A longitudinal analysis of student creativity scripts. *Journal of Creative Behavior*, *42*, 237–254. doi:10.1002/j.2162-6057.2008.tb01298.x

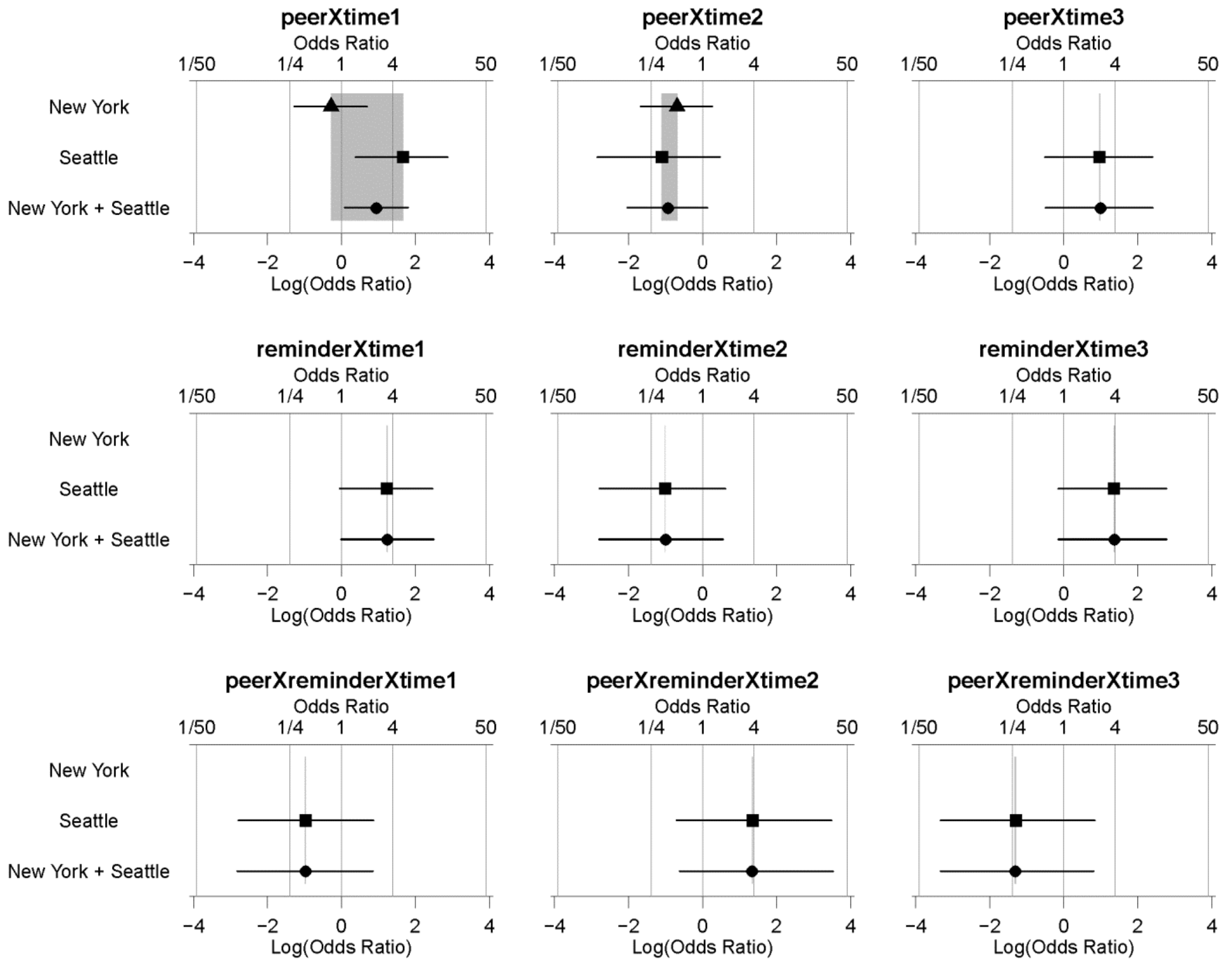

Figure 1. Multi-stage Bayesian Multilevel Logistic Regression Estimates of New York and Seattle studies

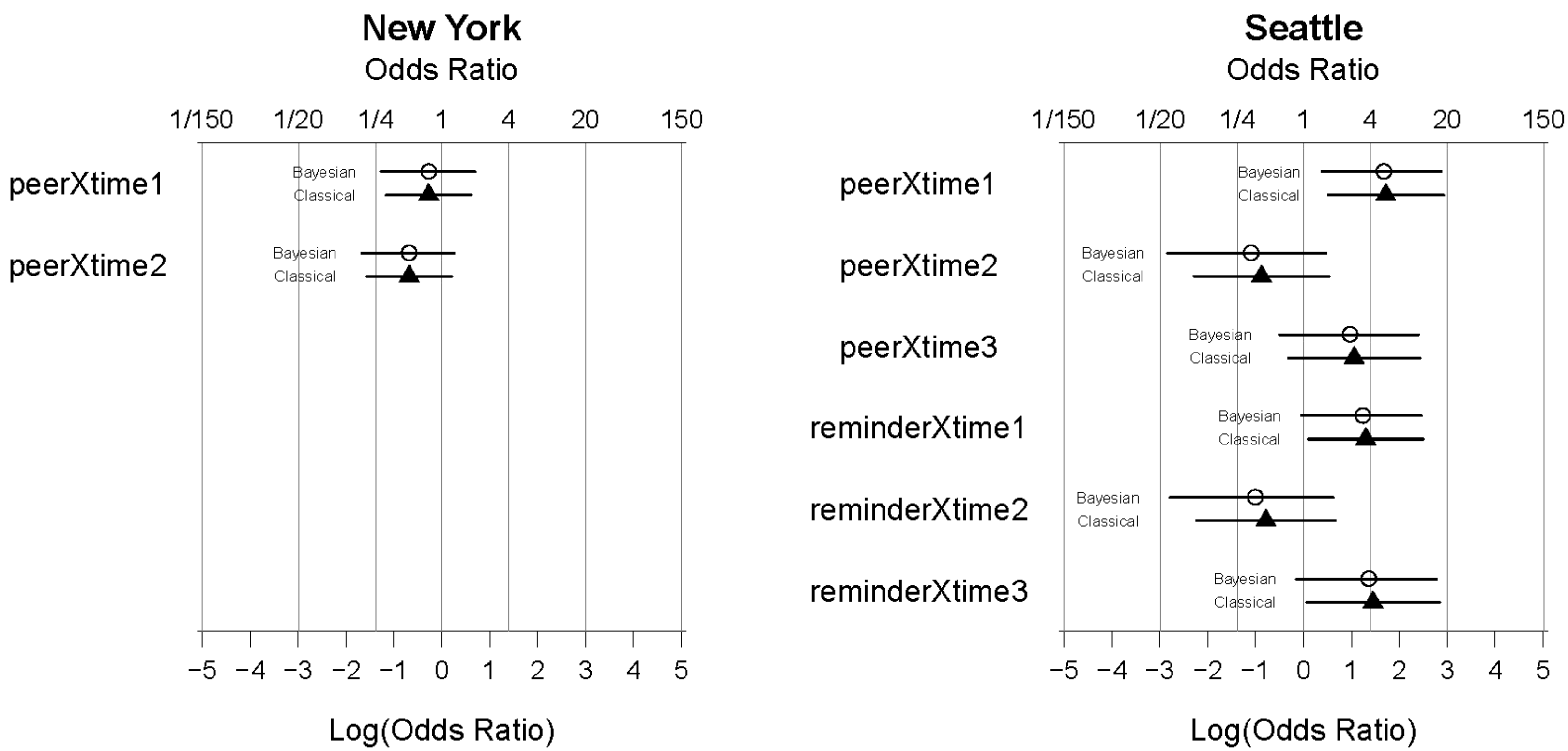

Figure 2. Multi-stage Bayesian Multilevel Linear Regression Estimates of New York, Seattle, and Beijing studies

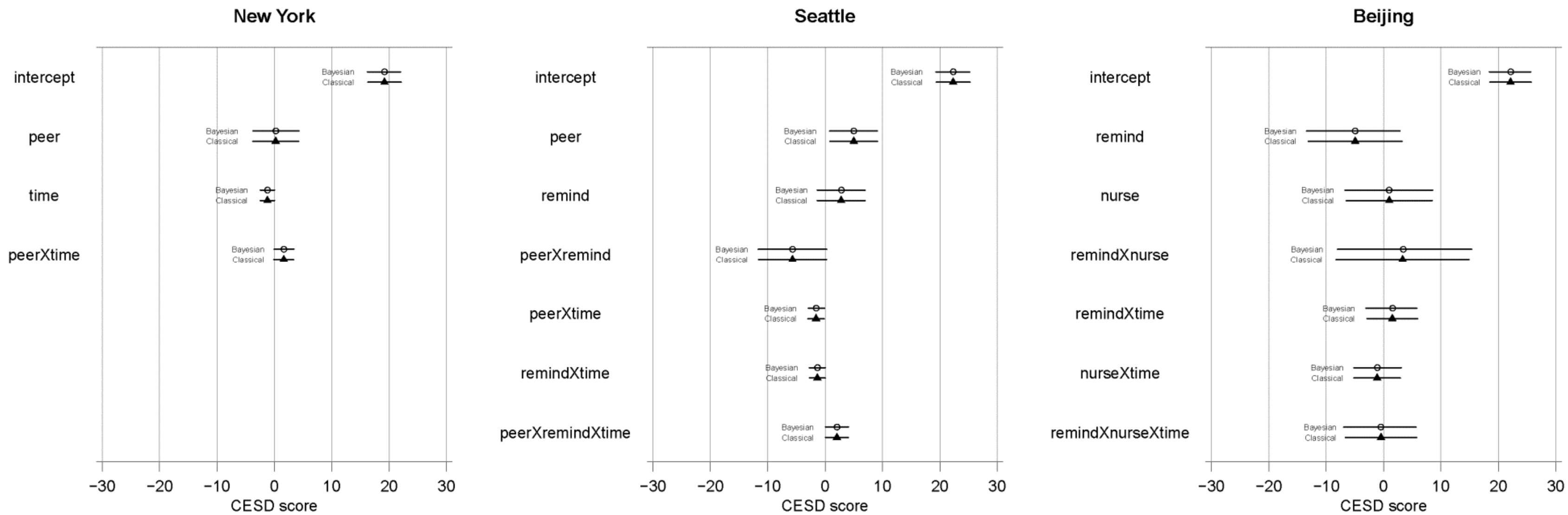

Figure 3. Sensitivity of Bayesian versus Classical Estimates of Multilevel Logistic Regression Analyses

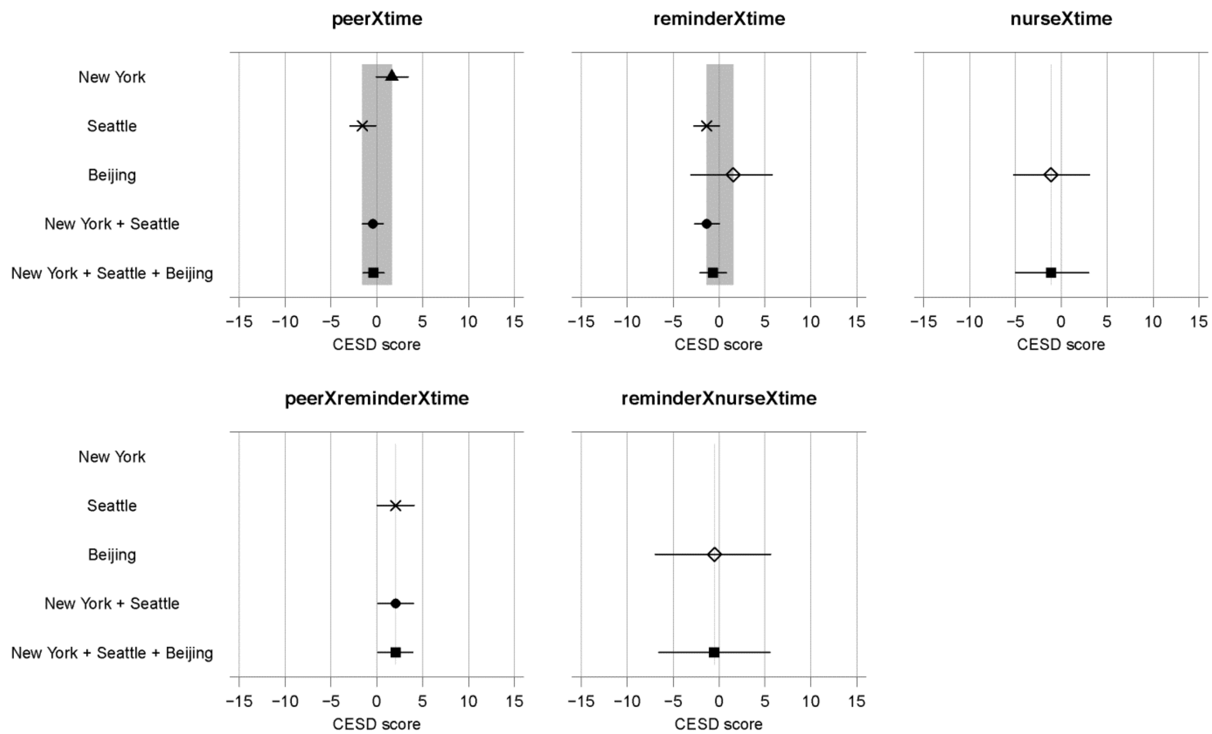

Figure 4. Sensitivity of Bayesian versus Classical Estimates of Multilevel Linear Regression Analyses

Table 1. Bayesian Population-Averaged Growth Curve Model in HLM Notation Evaluating the Peer and Reminder Interventions with Medication Adherence Data from New York and Seattle

Level I model:

$$\mathrm{logit}(\mathrm{ADHERE}_{ti}) = \pi_{0i} + \pi_{1i}(\mathrm{TIME1}_{ti}) + \pi_{2i}(\mathrm{TIME2}_{ti}) + \pi_{3i}(\mathrm{TIME3}_{ti}) + e_{ti},$$

where t = time point, i = participant, and $e_{ti} \sim \mathrm{N}(0, \sigma_e^2)$

Level II model:

$$\begin{aligned}
\pi_{0i} = {} & \boldsymbol{\beta_{00}} + \beta_{01}(\mathrm{SEA}_i) + \\
& \boldsymbol{\beta_{02}(\mathrm{PEER}_i \times \mathrm{NY}_i)} + \beta_{03}(PEER_i \times SEA_i) + \\
& \beta_{04}(REMINDER_i \times SEA_i) + \\
& \beta_{05}(PEER_i \times REMINDER_i \times SEA_i)
\end{aligned}$$

$$\begin{aligned}
\pi_{1i} = {} & \boldsymbol{\beta_{06}(\mathrm{NY}_i)} + \beta_{07}(\mathrm{SEA}_i) + \\
& \boldsymbol{\beta_{08}(\mathrm{PEER}_i \times \mathrm{NY}_i)} + \beta_{09}(\mathrm{PEER}_i \times \mathrm{SEA}_i) + \\
& \beta_{10}(\mathrm{REMINDER}_i \times \mathrm{SEA}_i) + \\
& \beta_{11}(\mathrm{PEER}_i \times \mathrm{REMINDER}_i \times \mathrm{SEA}_i)
\end{aligned}$$

$$\begin{aligned}
\pi_{2i} = {} & \boldsymbol{\beta_{12}(\mathrm{NY}_i)} + \beta_{13}(\mathrm{SEA}_i) + \\
& \boldsymbol{\beta_{14}(\mathrm{PEER}_i \times \mathrm{NY}_i)} + \beta_{15}(\mathrm{PEER}_i \times \mathrm{SEA}_i) + \\
& \beta_{16}(\mathrm{REMINDER}_i \times \mathrm{SEA}_i) + \\
& \beta_{17}(\mathrm{PEER}_i \times \mathrm{REMINDER}_i \times \mathrm{SEA}_i)
\end{aligned}$$

$$\begin{aligned}
\pi_{3i} = {} & \beta_{18}(\mathrm{SEA}_i) + \\
& \beta_{19}(\mathrm{PEER}_i \times \mathrm{SEA}_i) + \\
& \beta_{20}(\mathrm{REMINDER}_i \times \mathrm{SEA}_i) + \\
& \beta_{21}(\mathrm{PEER}_i \times \mathrm{REMINDER}_i \times \mathrm{SEA}_i)
\end{aligned}$$

Generalized linear mixed model:

$\mathrm{logit}(\mathrm{ADHERE}_{ti}) = \boldsymbol{\beta_{00}} + \beta_{01}(\mathrm{SEA}_{ti}) +$	[Intercept]
$\boldsymbol{\beta_{02}(\mathrm{PEER}_{ti} \times \mathrm{NY}_{ti})} + \beta_{03}(\mathrm{PEER}_{ti} \times \mathrm{SEA}_{ti}) +$	[Peer]
$\beta_{04}(\mathrm{REMINDER}_{ti} \times \mathrm{SEA}_{ti}) +$	[Reminder]
$\beta_{05}(\mathrm{PEER}_{ti} \times \mathrm{REMINDER}_{ti} \times \mathrm{SEA}_{ti}) +$	[Peer × Reminder]
$\boldsymbol{\beta_{06}(\mathrm{TIME1}_{ti} \times \mathrm{NY}_{ti})} + \beta_{07}(\mathrm{TIME1}_{ti} \times \mathrm{SEA}_{ti}) +$	[Time1)
$\boldsymbol{\beta_{08}(\mathrm{PEER}_{ti} \times \mathrm{TIME1}_{ti} \times \mathrm{NY}_{ti})} + \beta_{09}(\mathrm{PEER}_{ti} \times \mathrm{TIME1}_{ti} \times \mathrm{SEA}_{ti}) +$	[Peer × Time1]
$\beta_{10}(\mathrm{REMINDER}_{ti} \times \mathrm{TIME1}_{ti} \times \mathrm{SEA}_{ti}) +$	[Reminder × Time1]
$\beta_{11}(\mathrm{PEER}_{ti} \times \mathrm{REMINDER}_{ti} \times \mathrm{TIME1}_{ti} \times \mathrm{SEA}_{ti}) +$	[Peer × Reminder × Time1]
$\boldsymbol{\beta_{12}(\mathrm{TIME2}_{ti} \times \mathrm{NY}_{ti})} + \beta_{13}(\mathrm{TIME2}_{ti} \times \mathrm{SEA}_{ti}) +$	[Time2]
$\boldsymbol{\beta_{14}(\mathrm{PEER}_{ti} \times \mathrm{TIME2}_{ti} \times \mathrm{NY}_{ti})} + \beta_{15}(\mathrm{PEER}_{ti} \times \mathrm{TIME2}_{ti} \times \mathrm{SEA}_{ti}) +$	[Peer × Time2]
$\beta_{16}(\mathrm{REMINDER}_{ti} \times \mathrm{TIME2}_{ti} \times \mathrm{SEA}_{ti}) +$	[Reminder × Time2]
$\beta_{17}(\mathrm{PEER}_{ti} \times \mathrm{REMINDER}_{ti} \times \mathrm{TIME2}_{ti} \times \mathrm{SEA}_{ti}) +$	[Peer × Reminder × Time2]
$\beta_{18}(\mathrm{TIME3}_{ti} \times \mathrm{SEA}_{ti}) +$	[Time3]
$\beta_{19}(\mathrm{PEER}_{ti} \times \mathrm{TIME3}_{ti} \times \mathrm{SEA}_{ti}) +$	[Peer × Time3]
$\beta_{20}(\mathrm{REMINDER}_{ti} \times \mathrm{TIME3}_{ti} \times \mathrm{SEA}_{ti}) +$	[Reminder × Time3]
$\beta_{21}(\mathrm{PEER}_{ti} \times \mathrm{REMINDER}_{ti} \times \mathrm{TIME3}_{ti} \times \mathrm{SEA}_{ti}) + e_{ti}$	[Peer × Reminder × Time3]

Note. π_{0i} is the intercept (baseline) of the medication adherence trajectory for the ith participant, and π_{1i}, π_{2i}, π_{3i}, are the slopes (baseline versus 3, 6, and 9 months) of the medication adherence trajectories for the ith participant; Each row represents a model parameter with separate estimates for each study city.

Table 2. Bayesian Growth Curve Model in HLM notation Evaluating Peer, Reminder, and Nurse Interventions with Depression Data from New York, Seattle, and Beijing

Level I model:

$\mathrm{CESD}_{ti} = \pi_{0i} + \pi_{1i}(\mathrm{TIME}_{ti}) + e_{ti}$, where t = time point, i = participant and $e_{ti} \sim \mathrm{N}(0, \sigma_e^2)$

Level II model:

$$
\begin{aligned}
\pi_{0i} = {} & \boldsymbol{\beta_{00}} + \beta_{01}(\mathrm{SEA}_i) + \beta_{02}(\mathrm{BEI}_i) + \\
& \boldsymbol{\beta_{03}(\mathrm{PEER}_i \times \mathrm{NY}_i)} + \beta_{04}(\mathrm{PEER}_i \times \mathrm{SEA}_i) + \\
& \beta_{05}(\mathrm{REMINDER}_i \times \mathrm{SEA}_i) + \beta_{06}(\mathit{REMINDER}_i \times \mathit{BEI}_i) + \\
& \beta_{07}(\mathit{NURSE}_i \times \mathit{BEI}_i) + \\
& \beta_{08}(\mathrm{PEER}_i \times \mathrm{REMINDER}_i \times \mathrm{SEA}_i) + \\
& \beta_{09}(\mathit{NURSE}_i \times \mathit{REMINDER}_i \times \mathit{BEI}_i) + \\
& r_{0i},\ r_{0i} \sim \mathrm{N}(0, \sigma_0^2)
\end{aligned}
$$

$$
\begin{aligned}
\pi_{1i} = {} & \boldsymbol{\beta_{10}(\mathrm{NY}_i)} + \beta_{11}(\mathrm{SEA}_i) + \beta_{12}(\mathit{BEI}) + \\
& \boldsymbol{\beta_{13}(\mathrm{PEER}_i \times \mathrm{NY}_i)} + \beta_{14}(\mathrm{PEER}_i \times \mathrm{SEA}_i) + \\
& \beta_{15}(\mathrm{REMINDER}_i \times \mathrm{SEA}_i) + \beta_{16}(\mathit{REMINDER}_i \times \mathit{BEI}_i) + \\
& \beta_{17}(\mathit{NURSE}_i \times \mathit{BEI}_i) + \\
& \beta_{18}(\mathrm{PEER}_i \times \mathrm{REMINDER}_i \times \mathrm{SEA}_i) + \\
& \beta_{19}(\mathit{NURSE}_i \times \mathit{REMINDER}_i \times \mathit{BEI}_i)
\end{aligned}
$$

Linear mixed model:

$\mathrm{CESD}_{ti} = \boldsymbol{\beta_{00}} + \beta_{01}(\mathrm{SEA}_{ti}) + \beta_{02}(\mathit{BEI}_{ti}) +$	[Intercept]
$\boldsymbol{\beta_{03}(\mathrm{PEER}_{ti} \times \mathrm{NY}_{ti})} + \beta_{04}(\mathrm{PEER}_{ti} \times \mathrm{SEA}_{ti}) +$	[Peer]
$\beta_{05}(\mathrm{REMINDER}_{ti} \times \mathrm{SEA}_{ti}) + \beta_{06}(\mathit{REMINDER}_{ti} \times \mathit{BEI}_{ti}) +$	[Reminder]
$\beta_{07}(\mathit{NURSE}_{ti} \times \mathit{TIME}_{ti} \times \mathit{BEI}_{ti}) +$	[Nurse]
$\beta_{08}(\mathrm{PEER}_{ti} \times \mathrm{REMINDER}_{ti} \times \mathrm{SEA}_{ti}) +$	[Peer × Reminder]
$\beta_{09}(\mathit{NURSE}_{ti} \times \mathit{REMINDER}_{ti} \times \mathit{BEI}_{ti}) +$	[Nurse× Reminder]
$\boldsymbol{\beta_{10}(\mathrm{TIME}_{ti} \times \mathrm{NY}_{ti})} + \beta_{11}(\mathrm{TIME}_{ti} \times \mathrm{SEA}_{ti}) + \beta_{12}(\mathit{TIME}_{ti} \times \mathit{BEI}_{ti}) +$	[Time]
$\boldsymbol{\beta_{13}(\mathrm{PEER}_{ti} \times \mathrm{TIME}_{ti} \times \mathrm{NY}_{ti})} + \beta_{14}(\mathrm{PEER}_{ti} \times \mathrm{TIME}_{ti} \times \mathrm{SEA}_{ti}) +$	[Peer × Time]
$\beta_{15}(\mathrm{REMINDER}_{ti} \times \mathrm{TIME}_{ti} \times \mathrm{SEA}_{ti}) + \beta_{16}(\mathit{REMINDER}_{ti} \times \mathit{TIME}_{ti} \times \mathit{BEI}_{ti}) +$	[Reminder × Time]
$\beta_{17}(\mathit{NURSE}_{ti} \times \mathit{TIME}_{ti} \times \mathit{BEI}_{ti}) +$	[Nurse × Time]
$\beta_{18}(\mathrm{PEER}_{ti} \times \mathrm{REMINDER}_{ti} \times \mathrm{TIME}_{ti} \times \mathrm{SEA}_{ti}) +$	[Peer × Reminder × Time]
$\beta_{19}(\mathit{NURSE}_{ti} \times \mathit{REMINDER}_{ti} \times \mathit{TIME}_{ti} \times \mathit{BEI}_{ti})$	[Nurse× Reminder × Time]
$r_{0i} + e_{ti}$	

Note. π_{0i} is the intercept (baseline) of the depression trajectory for the ith participant, and π_{1i} is the slope of the depression trajectories for the ith participant. Each row represents a model parameter with separate estimates for each city.

Table 3. Pearson Correlations Among Month 0, 3, 6, and 9 Adherence Outcome with Means, Standard Deviations, Skew, and Kurtosis Statistics

	Month 3	Month 6	Month 9	*M*	*SD*	*Skew*	*Kurtosis*
New York							
Month 0	.36	.24	-	.78	.34	-1.45	3.76
Month 3		.45	-	.80	.34	-1.56	3.90
Month 6				.72	.41	-1.04	2.29
Seattle							
Month 0	.29	.20	-.05	.92	.20	-2.96	12.09
Month 3		.29	.23	.89	.26	-2.52	8.41
Month 6			.33	.81	.34	-1.65	4.06
Month 9				.84	.32	-1.97	5.40

Table 4. Pearson Correlations Among Month 0, 3, 6, and 9 Depressive Symptom Outcomes with Means, Standard deviations, Skew, and Kurtosis Statistics

	Month 3	Month 6	Month 9	*M*	*SD*	*Skew*	*Kurtosis*
New York							
Month 0	.72	.67	-	19.73	11.78	0.62	2.96
Month 3		.67	-	17.03	11.52	0.58	2.62
Month 6				19.61	13.03	0.69	2.89
Seattle							
Month 0	.54	.51	.51	25.24	11.55	0.12	2.06
Month 3		.64	.57	23.46	12.32	0.21	2.13
Month 6			.54	23.57	11.90	0.25	2.19
Month 9				23.09	12.56	0.24	2.21
Beijing							
Month 0	.53	.45	-	22.29	11.49	0.09	1.86
Month 3		.49	-	16.30	10.74	0.46	2.24
Month 6				14.97	10.44	0.57	2.27

Table 5. Summary of Bayesian Multilevel Logistic Regression Models Combining New York and Seattle Adherence Data

	New York only				Seattle only				New York + Seattle			
	B	*SD*	*95% CI*	*OR*	*B*	*SD*	*95% CI*	*OR*	*B*	*SD*	*95% CI*	*OR*
Intercept (Baseline status)												
Intercept (Standard of care)	**1.15**	**0.27**	**0.65 – 1.69**	-	**3.23**	**0.37**	**2.50 – 3.97**	-	**2.45**	**0.26**	**1.96 – 2.96**	-
Peer support	0.27	0.36	-0.45 – 0.97	1.31	-0.94	0.50	-1.92 – 0.06	0.39	-0.49	0.34	-1.18 – 0.18	0.61
Reminder					-0.91	0.53	-1.93 – 0.13	0.40	-0.92	0.53	-1.94 – 0.14	0.40
Peer × Reminder					0.91	0.75	-0.53 – 2.43	2.49	0.92	0.75	-0.49 – 2.45	2.52
Slope (Change since baseline)												
Time (Standard of care)												
- Month 3 vs. Baseline	0.26	0.36	-0.43 – 0.98	1.30	**-1.58**	**0.43**	**-2.41 – -0.73**	**0.21**	**-0.89**	**0.30**	**-1.47 – -0.30**	**0.41**
- Month 6 vs. Baseline	0.04	0.32	-0.58 – 0.67	1.04	-0.11	0.68	-1.42 – 1.36	0.89	-0.06	0.45	-0.91 – 0.88	0.94
- Month 9 vs. Baseline					-1.69	0.52	-2.73 – -0.69	0.18	-1.70	0.51	-2.72 – -0.69	0.18
Peer × Time												
- Month 3 vs. Baseline	-0.28	0.50	-1.27 – 0.69	0.76	**1.67**	**0.63**	**0.38 – 2.87**	**5.32**	**0.94**	**0.44**	**0.09 – 1.80**	**2.56**
- Month 6 vs. Baseline	-0.69	0.49	-1.67 – 0.26	0.50	-1.10	0.82	-2.84 – 0.47	0.33	-0.94	0.54	-2.04 – 0.13	0.39
- Month 9 vs. Baseline					0.97	0.74	-0.51 – 2.40	2.63	0.99	0.73	-0.49 – 2.40	2.69
Reminder × Time												
- Month 3 vs. Baseline					1.23	0.64	-0.05 – 2.47	3.44	1.24	0.64	-0.01 – 2.50	3.46
- Month 6 vs. Baseline					-1.01	0.85	-2.79 – 0.62	0.36	-1.01	0.85	-2.79 – 0.56	0.37
- Month 9 vs. Baseline					1.36	0.75	-0.15 – 2.78	3.89	1.37	0.74	-0.14 – 2.78	3.94
Peer × Reminder × Time												
- Month 3 vs. Baseline					-0.97	0.94	-2.78 – 0.87	0.38	-0.97	0.94	-2.81 – 0.86	0.38
- Month 6 vs. Baseline					1.35	1.06	-0.70 – 3.49	3.87	1.34	1.05	-0.61 – 3.53	3.82
- Month 9 vs. Baseline					-1.29	1.06	-3.33 – 0.84	0.27	-1.32	1.05	-3.33 – 0.81	0.27

Note. Estimates in bold possess a 95% Bayesian confidence interval that does not contain a zero effect.

Table 6. Bayesian Multilevel Linear Models Sequentially Combining Depression Data from New York, Seattle, and Beijing

	I. New York			II. New York + Seattle			III. New York + Seattle + Beijing		
	B	*SD*	*95% CI*	*B*	*SD*	*95% CI*	*B*	*SD*	*95% CI*
Intercept (Baseline status)									
Intercept	**19.17**	**1.48**	**16.21 – 21.98**	**20.85**	**0.99**	**18.87 – 22.74**	**21.28**	**0.96**	**19.45 – 23.23**
Peer	0.23	2.06	-3.72 – 4.28	**3.46**	**1.44**	**0.53 – 6.21**	**3.18**	**1.52**	**0.17 – 6.11**
Reminder				2.78	2.12	-1.24 – 7.01	0.97	1.93	-2.94 – 4.62
Nurse							1.01	4.19	-7.01 – 9.41
Peer × Reminder				-5.65	3.00	-11.58 – 0.19	-5.68	3.00	-11.57 – 0.20
Reminder × Nurse							3.26	6.47	-9.65 – 15.61
Slope (Change per assessment)									
Time (Standard of care)	-1.22	0.63	-2.44 – 0.02	-0.12	0.40	-0.93 – 0.64	**-0.78**	**0.38**	**-1.51 – -0.03**
Peer × Time	1.64	0.88	-0.05 – 3.40	-0.43	0.58	-1.58 – 0.70	-0.36	0.58	-1.48 – 0.78
Reminder × Time				-1.34	0.69	-2.66 – 0.05	-0.65	0.74	-2.07 – 0.80
Nurse × Time							-1.10	2.03	-4.97 – 2.99
Peer × Reminder × Time				**2.05**	**0.99**	**0.12 – 3.99**	**2.05**	**0.97**	**0.13 – 3.91**
Reminder × Nurse × Time							-0.52	3.11	-6.57 – 5.56
				Seattle only			Beijing only		
Intercept (Baseline status)									
Intercept (Standard of care)				**22.30**	**1.50**	**19.32 – 25.20**	**22.12**	**1.83**	**18.43 – 25.61**
Peer				**4.98**	**2.12**	**0.77 – 9.10**			
Reminder				2.82	2.13	-1.36 – 6.98	-4.98	4.18	-13.42 – 2.87
Nurse							0.94	3.88	-6.74 – 8.55
Peer × Reminder				-5.69	3.04	-11.62 – 0.31			
Reminder × Nurse							3.40	5.94	-8.04 – 15.33
Slope (Change per assessment)									
Time (Standard of care)				0.41	0.50	-0.58 – 1.39	**-3.77**	**1.05**	**-5.93 – -1.81**
Peer × Time				**-1.57**	**0.72**	**-2.92 – -0.11**			
Reminder × Time				-1.35	0.71	-2.74 – 0.04	1.54	2.27	-3.06 – 5.79
Nurse × Time							-1.12	2.10	-5.18 – 3.08
Peer × Reminder × Time				**2.06**	**1.02**	**0.07 – 4.07**			
Reminder × Nurse × Time							-0.50	3.21	-6.94 – 5.62

Note. Estimates in bold possess a 95% Bayesian confidence interval that does not contain a zero effect.

CURRICULUM VITAE

David Huh, Ph.D.
Department of Psychology
University of Washington
Box 351525
Seattle, WA 98195-1525

E-mail: dhuh@uw.edu
Phone: (206) 552-9848

EDUCATION

Doctorate
September 2006 – August 2012

University of Washington, Seattle, WA
Major Area: Clinical Psychology
Minor Area: Quantitative Psychology

Dissertation: "Developing a Bayesian Statistical Approach to Behavioral Intervention Trials"

Supervisory Committee: Jane M. Simoni, Ph.D., Brian P. Flaherty, Ph.D., John Miyamoto, Ph.D., Diane Morrison, Ph.D.

Master's
September 2006 – December 2008

University of Washington, Seattle, WA
Master of Science
Major Area: Clinical Psychology

Thesis: "A Comparison of Analysis of Covariance, Generalized Estimating Equations, and Generalized Linear Mixed Regression for Evaluating HIV Medication Adherence Interventions"

Advisors: Jane M. Simoni, Ph.D., Brian P. Flaherty, Ph.D.

Bachelor's
September 1999 – June 2003

University of Oregon, Eugene, OR
Bachelor of Science
Major Areas: Psychology, General Science

Thesis: "Voluntary versus involuntary immigration: The adjustment of Asian Pacific Americans in the context of family"

Honors: magna cum laude, Dean's list, departmental honors (psychology)

Advisor: Gordon C. Nagayama Hall, Ph.D.

September 2000 – December 2001

Northwestern University, Evanston, IL
Major: Psychology

Advisor: William Revelle, Ph.D.

GRANTS AWARDED

Training Supplement
May 2004 – January 2006

Test of a Dissonance Eating Disorder Prevention Program
National Institute of Mental Health (Minority Training Supplement to R01MH061957, Total costs = $101,039)

Principle Investigator/Mentor: Eric Stice, Ph.D.

RESEARCH EXPERIENCE

July 2009 – June 2011

Statistician, Indigenous Wellness Research Institute, University of Washington
Projects: Honor Project, Healthy Hearts Across Generations
Supervisors: Cynthia Pearson, Ph.D., Karina Walters, Ph.D.

September 2006 – June 2011

Research Assistant, Department of Psychology, University of Washington
Projects: Promoting Adherence for Life; Developing an Antiretroviral Adherence Program in China; Multi-Site Study for Adherence, Drug Resistance and Virologic and Clinical Outcomes; Addressing Depression and ART Adherence in HIV+ Latinos on the US-Mexico Border
Supervisors: Jane Simoni, Ph.D., Cynthia Pearson, Ph.D.

June – October 2006

Data Analyst, Abacus Consulting, LLC, Eugene, OR
Conducted statistical programming in SPSS to generate statistics on state-level standardized testing data.
Supervisors: Jeffrey Gau, M.S., John Seeley, Ph.D.

July 2004 – July 2006

Research Assistant, Oregon Research Institute and University of Texas at Austin
Projects: Body Project, Austin Adolescent Development Study
Maintained and statistically evaluated longitudinal data from ongoing etiologic and prevention studies relating to obesity and eating disorders among adolescent females.
Supervisor: Eric Stice, Ph.D.

October 2003 – June 2004

Interviewer, Oregon Survey Research Laboratory, Eugene, OR
Conducted interviews regarding substance use and sexual risk-taking among college students, social capital in eight Northwest states, and studies of mental and physical health among senior citizens.
Supervisors: Woody Carter, Ph.D., Robert Choquette, Ph.D.

June – August 2003

Intern, Latino Youth and Family Empowerment Project, Oregon Social Learning Center, Eugene, OR
Managed data and trained in ethnic minority research and culturally-informed research methods.
Supervisors: Charles Martinez, Ph.D., Betsy Ruth, M.S.W.

April – August 2002

Research Assistant, Early Intervention Foster Care Project, Oregon Social Learning Center, Eugene, OR
Conducted home interviews with parents of at-risk pre-school age foster children.
Supervisor: Odile Stout

PUBLICATIONS

Peer-Reviewed Journal Articles (Published and In Press)

10. Simoni, J. M., **Huh, D.,** Wilson, I., Goggin, K., Reynolds, N., Remien, R. H., et. al. (in press). Racial/ethnic disparities in ART adherence in the United States: Findings from the MACH14 study. *Journal of Acquired Immune Deficiency Syndromes*. doi: 10.1097/QAI.0b013e31825db0bd

9. Simoni, J. M., Yard, S. S., & **Huh, D.** (2012). Prospective prediction of viral suppression and immune response nine months after ART initiation in Seattle, WA [epub ahead of print]. *AIDS Care.* Retrieved from http://dx.doi.org/10.1080/09540121.2012.687821

8. **Huh, D.**, Flaherty, B. P., & Simoni, J. M. (2012). Optimizing the analysis of adherence interventions using logistic generalized estimating equations. *AIDS and Behavior, 16,* 422-431. doi: 10.1007/s10461-011-9955-5

7. **Huh, D.**, Stice, E., Shaw, H., & Boutelle, K. (2012). Female overweight and obesity in adolescence: Developmental trends and ethnic differences in prevalence, incidence, and remission. *Journal of Youth and Adolescence, 41,* 76-85. doi: 10.1007/s10964-011-9664-4

6. Yard, S. S., **Huh, D.**, King, K., Simoni, J. M. (2011). Patient-level moderators of the efficacy of peer support and pager reminder interventions to promote antiretroviral adherence. *AIDS and Behavior, 15,* 1596-1604. doi: 10.1007/s10461-011-0001-4

5. Lehavot, K., **Huh, D.**, Walters, K. L., King, K. M., Andrasik, M. P., & Simoni, J. M. (2011). Buffering effects of general and medication-specific social support on the association between substance use and HIV medication adherence. *AIDS Patient Care and STDs, 25,* 181-189. doi: 10.1089/apc.2010.0314

4. Simoni, J. M., Chen, W., **Huh, D.**, Fredriksen-Goldsen, K. I., Pearson, C., Zhao, H., et al. (2011). A preliminary randomized controlled trial of a nurse-delivered medication adherence intervention among HIV-positive outpatients initiating antiretroviral therapy in Beijing, China. *AIDS and Behavior, 15*, 919-929. doi: 10.1007/s10461-010-9828-3

3. Harris, L. T., Lehavot, K., **Huh, D.**, Yard, S., Andrasik, M. P., Dunbar, P. J., et al. (2010). Two-way text messaging for health behavior change among human immunodeficiency virus-positive individuals. *Telemedicine and e-Health, 16,* 1024-2029. doi: 10.1089/tmj.2010.0050

2. Simoni, J. M., **Huh, D**., Frick, P., Pearson, C. R., Andrasik, M. P., & Hooton, M. (2009). Peer support and pager messaging to promote antiretroviral modifying therapy in Seattle: A randomized controlled trial. *Journal of Acquired Immune Deficiency Syndromes, 52,* 465-473. doi: 10.1097/QAI.0b013e3181b9300c

1. **Huh, D.**, Tristan, J., Wade, E., & Stice, E. (2006). Does problem behavior elicit poor parenting? A prospective study of adolescent girls. *Journal of Adolescent Research, 21,* 185-204. doi: 10.1177/0743558405285462

Book Chapters

1. Walters, K. L., Beltran, R., **Huh, D.**, & Evans-Campbell, T. (2011). Dis-placement and dis-ease: Land, place, and health among American Indians and Alaska Natives. In D. Takeuchi (Ed.), *Expanding the Boundaries of Place (pp. 163-199).* New York, NY: Springer. doi: 10.1007/978-1-4419-7482-2_10

PRESENTATIONS

11. Simoni, J. M., Wiebe, J. S., Sauceda, J. A., **Huh, D.**, Longoria, V., Sanchez, G., et al. (2012, June). A culturally adapted intervention to treat depression and ART nonadherence on the U.S.-Mexico border: Final (promising!) results from a pilot RCT. Oral presentation at the 7th NIMH/IAPAC International Conference on HIV Treatment Adherence, Miami, FL.

10. Pantalone, D. W., **Huh, D.**, Nelson, K., Pearson, C. R., & Simoni, J. M. (2011, November). Prospective predictors of unprotected sex in HIV-positive sexual minority men starting or re-starting antiretroviral therapy. In J. T. Parsons (Chair), *HIV-Related Sexual Risk among MSM.* Symposium presentation at the 51st Annual Meeting of the Society for the Scientific Study of Sexuality, Houston, TX.

9. Simoni, J. M., Wiebe, J. S., Sauceda, J. A., Sanchez, G., Longoria, V., **Huh, D.**, et al. (2011, May). *A culturally adapted intervention to treat depression and ART nonadherence on the*

U.S.-Mexico border: Feasibility and initial results from a pilot RCT. Oral presentation at the 6th NIMH/IAPAC International Conference on HIV Treatment Adherence, Miami, FL.

8. Simoni, J. M., Wilson, I., Goggin, K., Reynolds, N., Remien, R. H., **Huh, D.**, et. al. (2011, May). Can substance use and depression account for race/ethnicity disparities in ART adherence? A closer look at findings from the MACH14 study. Oral presentation at the 6th NIMH/IAPAC International Conference on HIV Treatment Adherence, Miami, FL.

7. Simoni, J. M., Wiebe, J. S., Sauceda, J.A., Sanchez, G., Longoria, V., **Huh, D.**, et al. (2011, May). *A culturally adapted intervention to treat depression and ART nonadherence on the U.S.-Mexico border: Feasibility and initial results from a pilot RCT.* Oral presentation at the 6th NIMH/IAPAC International Conference on HIV Treatment Adherence, Miami, FL.

6. **Huh, D**., Simoni, J. M., & Flaherty, B. P. (2009, June). *A comparison of ANCOVA, generalized estimating equations, and generalized linear mixed regression for evaluating HIV medication adherence interventions.* Oral presentation at the Western North American Region/International Biometric Society Annual Meeting, Portland, OR.

5. Simoni, J. M., **Huh, D.**, Chen, W., & Nelson, K. (2009, April). *Promising results from an RCT of a nurse-delivered ART adherence promotion program in China.* Paper presented at the 4th NIMH/IAPAC International Conference on HIV Treatment Adherence, Miami, FL.

4. Simoni, J. M., **Huh, D**., Frick, P. A., Pearson, C. R., Lehavot, K., Dunbar, P. J., et al. (2009, April). *Final results from an RCT evaluating the effect of peer and pager support on antiretroviral adherence.* Paper presented at 4th NIMH/IAPAC International Conference on HIV Treatment Adherence, Miami, FL.

3. Simoni, J. M., Chen, W., **Huh, D.**, Zhao, H., Fredriksen-Goldsen, K. I., Pearson, C., et al. (2009, April). *An RCT of a nurse-delivered medication adherence intervention among antiretroviral therapy-naïve outpatients in Beijing, China.* Presented at 4th International Conference on HIV Treatment Adherence, Miami, FL.

2. Simoni, J., Frick, P. A., Pearson, C., Hooton, T., **Huh, D.**, & Rojas, S. (2007, March). *Preliminary RCT findings demonstrate effect of peer and pager support on antiretroviral adherence.* Paper presented at 2nd International Conference on HIV Treatment Adherence, Jersey City, NJ.

1. Alonso, T., Decker, R., & **Huh, D.** (2003, July). *Parental involvement on Latino youth outcomes*. Oral presentation the Annual NIDA Summer Internship Conference.

Invited Talks

3. **Huh, D.** (2007, December). *Findings from Project PAL: Did peer and pager support promote antiretroviral adherence at Madison Clinic?* Presentation at Madison Clinic, Harborview Medical Center, Seattle, WA.

2. Simoni, J. M., **Huh, D.**, & Pearson, C. R. (2007, December). *Findings from Project PAL: Did peer and pager support promote antiretroviral adherence at Madison Clinic?* Paper presented at University of Washington CFAR Research Day, Harborview Medical Center, Seattle, WA.

1. Simoni, J. M., & **Huh, D.** (2007, June). *Findings from Project PAL: Did peer and pager support promote antiretroviral adherence at Madison Clinic?* Presentation at 2007 AIDS & STD Research Symposium, Harborview Medical Center, Seattle, WA.

Posters

6. **Huh, D.**, Simoni, J., & Flaherty, B. (2011, May). A comparison of logistic generalized estimating equations with planned contrasts versus growth curves for evaluating antiretroviral adherence interventions. Poster presented at the 6th NIMH/IAPAC International Conference on HIV Treatment Adherence, Miami, FL.

5. Simoni, J. M., **Huh, D.**, Frick, P., & Hooton, T. (2009, August). *Do adherence interventions decrease HIV viral load? A moderator analysis.* Poster presented at the 117th Annual Convention of the American Psychological Association, Boston, MA.

4. Lehavot, K., **Huh, D.**, Simoni, J. M., Andrasik, M. P., King, K. M., & Walters, K. L. (2009, August). *Social support moderates the substance use-medication adherence relationship among HIV+ individuals?* Poster presented at the 117th Annual Convention of the American Psychological Association, Boston, MA.

3. Simoni, J. M., **Huh, D.**, Frick, P., & Hooton, T. (2009, August). *Do adherence interventions decrease HIV viral load? A moderator analysis.* Poster presented at the 117th Annual Convention of the American Psychological Association, Boston, MA.

2. Simoni, J. M., **Huh, D.**, Yard, S., Balsam, K., Lehavot, K., & Pearson, C. (2008, August). *Does peer support improve mental health outcomes? Results from an RCT among HIV-positive men and women.* Poster presented at the 116th Annual Convention of the American Psychological Association, Boston, MA.

1. **Huh, D.**, Simoni, J., & Flaherty, B. (2008, March). *Statistical methods for randomized controlled trials: A comparison of ANCOVA, linear mixed regression, and generalized estimating equations for evaluating HIV medication adherence interventions*. Poster presented at the 3rd NIMH/IAPAC International Conference on HIV Treatment Adherence, Jersey City, NJ.

CLINICAL EXPERIENCE

July 2011 – June 2012

Psychology Resident, Department of Psychiatry and Behavioral Sciences, University of Washington.
Mentors: Mary Larimer, Ph.D., David Atkins, Ph.D.

October 2009 – November 2010	**Psychology Extern,** Inpatient Psychiatry, University of Washington School of Medicine. Provided group and individual evidence-based interventions with voluntary psychiatric inpatients. *Supervisors:* Steven Vannoy, Ph.D., Mary Larimer, Ph.D.
March – August 2009	**Psychology Extern,** Rehabilitation Psychology, University of Washington School of Medicine. Conducted background interviews with out-patients in the Traumatic Brain Injury Clinic. Responded to consult requests from outside departments. Provided brief evidence-based interventions with rehabilitation medicine inpatients as part of an interdisciplinary team. *Supervisors:* Jeanne Hoffman, Ph.D., Mary Pepping, Ph.D.
March – August 2009	**Family Therapy Team Therapist,** University of Washington Psychological Services and Training Clinic. Was a member of a team of 5 therapists providing consultation from behind-a-mirror to the co-leading therapists. Case involved a family of three generations and utilized family systems and cognitive-behavioral techniques. *Supervisor:* Corey Fagan, Ph.D.
November 2007 – June 2009	**Graduate Student Therapist,** University of Washington Psychological Services and Training Clinic. Provided evidence-based outpatient psychotherapy to three individual clients. *Supervisors:* Corey Fagan, Ph.D., Charles Huffine, M.D., Neal Teng, Ph.D., Fransing Daisy, Ph.D.

HONORS AND AWARDS

June 2009	Travel Award, Western North American Region/International Biometric Society Annual Meeting
August 2008	Wagner Memorial Travel Award, Department of Psychology, University of Washington

PROFESSIONAL AFFILIATIONS AND SERVICE

Organizational Memberships

American Psychological Association	Division 5: Evaluation, Measurement, and Statistics Division 38: Health Psychology

International Biometric Society Western North American Region

Committee Memberships

September 2009 – June 2011 Diversity Steering Committee, Department of Psychology, University of Washington, *Graduate Student Representative*

September 2009 – June 2011 Graduate Training Committee, Department of Psychology, University of Washington, *Graduate Student Representative*

LANGUAGES

Spanish I have conversational proficiency in standard and Rioplatense Spanish.

PROFESSIONAL REFERENCES

(Available upon request)