BAYESIAN ANALYSIS OF LATENT TRAIT HIERARCHICAL MODELS FOR MULTIPLE BINARY OUTCOMES IN CLUSTER RANDOMIZED CLINICAL TRIALS

by

Xinhua Zhao

B. Med, Shanxi Medical University, Taiyuan, China, 1994

M. Med, Chinese Academy of Preventive Medicine, Beijing, China, 1997

M.S, University of Georgia, 2001

Submitted to the Graduate Faculty of

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2010

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This dissertation was presented

by

Xinhua Zhao

It was defended on

September 15th, 2010

and approved by

Dissertation Advisor:
Roslyn A. Stone, PhD
Associate Professor
Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

Committee Member:
Howard E. Rockette, PhD
Professor
Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

Sati Mazumdar, PhD
Professor
Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

Feifei Ye, PhD
Assistant Professor
Psychology in Education, School of Education, University of Pittsburgh

Michael J. Fine, MD
Professor
Department of Medicine, School of Medicine, University of Pittsburgh

BAYESIAN ANALYSIS OF LATENT TRAIT HIERARCHICAL MODELS FOR MULTIPLE BINARY OUTCOMES IN CLUSTER RANDOMIZED CLINICAL TRIALS

Xinhua Zhao, PhD

University of Pittsburgh, 2010

In clinical trials, multiple endpoints for treatment efficacy often are obtained, and in addition, data may be collected hierarchically. Statistical analyses become very challenging for this multidimensional hierarchical data, particularly with data collected at more than two levels. We propose a latent variable approach to assess an intervention effect on multiple binary outcomes from three-level hierarchical data. This approach incorporates the correlation structure into one or more latent outcomes, and simultaneously regresses the latent outcome(s) on observed covariates. Random effects are included to model the hierarchical structure. Parameters estimation is done using a fully Bayesian approach implemented in WinBUGS.

We first illustrate the approach in a cluster randomized clinical trial of three interventions to improve the processes of care for outpatients with pneumonia. Four binary outcomes are collected at the patient-level and clustered at the provider and clinic site levels. Simulation studies are conducted to check the algorithm and computational implementation. Then, we extend the one latent trait model to a two-latent trait model using eight outcomes from both outpatient and inpatient care. This latent modeling approach provides a comprehensive way to analyze multivariate hierarchical data. The method not only allows assessment of intervention effects with respect to multiple outcomes, but also assesses the relationship between outcomes, identifies those outcomes that carry the most information about the latent trait(s), and provides a summary measure of the "quality of care" at each clinical site.

This work extends existing methods to model multivariate binary endpoints in a cluster-randomized clinical trial. The public health significance of this study is the potential usefulness of this approach to quantify intervention (or exposure) effects with regard to multiple outcomes in hierarchical data setting, which arises frequently in medical and epidemiologic studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

First, I would like thank my advisor, Dr. Roslyn A. Stone for introducing this interesting topic to me, mentoring me through the course of my dissertation work with insightful guidance and persistent encouragement. I am truly grateful to Dr. Stone's tremendous patience in revision of my proposal and this dissertation. Without her support, I would not have completed this work.

Next, I would like to thank my committee members: Dr. Howard E. Rockette, Dr. Sati Mazumdar, Dr. Feifei Ye and Dr. Michael J. Fine, for their valuable suggestions and thoughtful comments on the statistical and practical aspects of my research. Special thanks to Dr. Feifei Ye for valuable discussions on the model specification and programming issues, Dr. Howard E. Rockette and Dr. Sati Mazumdar for their wonderful coursework, and Dr. Michael J. Fine for providing me with the wonderful and rich data that motivated the main idea for my dissertation.

As a part-time PhD student, I also would like to thank University of Pittsburgh for the education benefit, and thank my colleagues and supervisors from the Epidemiological Data Center (2002-2007) and the Center for Health Equity Research and Promotion (CHERP, 2008-present) for their understanding and support.

Finally, I dedicate this dissertation to my husband and my two children for their love and support.

1.0 INTRODUCTION

There are many situations in medical and other applied research settings where the outcomes of interest cannot be characterized by one single measurement on the individuals under study. For example, no single measure exists for outcomes such as blood pressure, quality of life, or physical functioning. To effectively capture all aspects of such outcomes, a number of measurements commonly are used. For example, blood pressure can be measured by systolic and diastolic blood pressures. Quality of life often is measured with a multi-item self-report questionnaire. Physical functioning is measured using an extensive self-report questionnaire and/or multiple performance-based tests. The result of such measurements is the representation of blood pressure, quality of life, or physical functioning by a set of scores for each individual.

1.1 STATEMENT OF THE PROBLEM

In clinical trials, multiple endpoints for treatment efficacy often are obtained (Pocock, Geller, and Tsiatis, 1987). For example, in stroke recovery no single outcome can serve as a gold standard indicative of treatment efficacy. The primary objective of analyses of this kind is not to identify the particular outcomes that differ between the groups, but rather to use all the data at hand to establish whether there is a difference between the groups. Several statistical issues arise in evaluating a treatment effect from this type of data. One main issue is the multidimensionality.

Each of the several response variables measures a slightly different aspect of the response. because no single response variable suffices as a main outcome variable, methods that collectively consider all of the response variables are desired. The main question is how to use the multiple response variables to obtain a summary measure of the treatment effect that is readily interpretable. The problem is challenging because the multiple response variables may be defined on different numerical scales. If the data are collected hierarchically, the result is multidimensional hierarchical data. These methods also must account for the autocorrelation between individuals within the same cluster within each response variable, and the cross-correlation between different response variables both across clusters and within the same cluster. Frequently, from a substantive perspective, there is a need to summarize all the multidimensional measures into a unidimensional composite score, such as an overall measure of quality of care in health care research.

The frequently used approaches include individual outcome data analysis, dimension reduction of the data, and global test procedures. Each of those approaches fails to characterize the relationship between outcomes or summarize the variables. Latent variable models provide a natural way to analyze the complex multivariate data in this setting. A latent variable model is any model that includes the unobserved random variables. These models have been employed extensively in the areas of psychological and educational testing (Baker F.B. 1992) and the social sciences (Eye and Clogg, 1994). In recent years, the utility of latent variable models has been recognized and the use is increasing in medical research (for example see Catalano and Ryan, 1992; Bollen and Long, 1993; Legler and Ryan, 1997; Sammel, Ryan, and Legler, 1997). However, latent variable models seldom have been used to test hypotheses about clinical outcomes in clinical trials and other designed studies (Donaldson 2003).

The overwhelming majority of the literature on latent variable models is frequentist in nature, using maximum likelihood estimation. MLE commonly treats latent variables as random and parameters as fixed. Inference usually is based on the marginal likelihood, the likelihood of the data given the latent variables, integrated over the latent variable distribution. The EM algorithm often is applied to maximize the likelihood. Unfortunately, in general no closed form for the multi-dimensional integrals exists, so that some approximations are required. Also, extra effort such as calculating the observed information matrix from the Louis formula is required to estimate standard errors (Louis, 1982).

Bayesian estimation of latent variable models using Markov Chain Monte Carlo (MCMC) is an attractive alternative to maximum likelihood. Bayesians treat both latent variables and parameters as random variables. The difficulties induced by the complexities of the model and the multi-dimensional integrals can be handled efficiently using powerful computing tools such as the Gibbs sampler. An additional benefit is that samples are available from the joint posterior distribution of the latent variables. Often, these samples can be used to obtain important insights into structural relationships, which may not be apparent from the parameter estimates, such as checking the normality of the posteriors samples of latent variable to capture lack of model fit (Sik-Yum Lee 2007). Since about 2000, a number of authors have used MCMC methods to implement Bayesian analysis in various structural equation models (SEMs), involving nonlinear structure, heterogeneity and multilevel data (Lee and Song 2004, Lee and Song 2003, Dunson, et al. 2000). The majority of the Bayesian approaches for complex SEMs were developed under the crucial assumption that the conditional distribution of the manifest (observed) variables, given the latent variables, is normal, while other distributions, such as the binomial (unordered binary data), have received limited consideration.

Very recently, latent variable modeling approaches have been used to profile health care providers on quality of health care. Quality of healthcare is an abstract and multidimensional construct that cannot be measured directly. There are three dimensions of quality of care: structure, process, and outcome. Structure measures are characteristics of the health provider. Process measures are the components of the encounter between a physician or other healthcare professional and a patient. Outcome measures refer to the patient's subsequent health status (Blumenthal 1996). Normand et al. published a series of articles using latent variable models to quantify quality of health care using Gibbs sampling, involving a 2-parameter Normal-Ogive model (multivariate probit model) for multiple cross-sectional binary outcomes (2000, 2008), and multilevel multidimensional latent variable models under multivariate normal distribution with a threshold model for mixed binary and continuous outcomes in cross-sectional (2003) and longitudinal (2006) data settings. The aim of these studies was to estimate the unidimensional latent score in order to profile health care providers.

Latent variable models have played little role in analyzing clinical trial data. For binary outcomes, item response theory (IRT) models with a probit-link were commonly used in literature. Given the easy interpretation and popularity of logit-link in clinical data settings, we will use logit link for the proposed method, which generalizes the conditional distribution of the manifest variable given the latent variable from the normal distribution to other exponential family distributions. In addition, existing standard software cannot analyze multiple outcomes from hierarchical data with more than two levels.

In our work, we propose a general latent variable model to analyze an intervention effect with regard to multiple binary outcomes for a three-level model with multiple outcomes at the patient-level and clustering at the provider and site levels, using MCMC Bayesian estimation.

This latent modeling approach provides a comprehensive way to analyze multivariate hierarchical data. Not only does the method allow assessment of intervention effects with respect to multiple outcomes, but it also quantifies the relationship between outcomes, identifies those outcomes carry the most information about the latent trait, and provides a summary measure of the "quality of care" of each clinical site.

1.2 MOTIVATING EXAMPLE: THE EDCAP STUDY

The Emergency Department Community Acquired Pneumonia (EDCAP) Trial (Yealy, Auble, Stone et al., 2004) motivated this proposed method. The EDCAP study was designed to compare the effectiveness and safety of three guideline implementation strategies of increasing intensity (low-intensity, moderate-intensity, and high-intensity) on quality of care of patients with community acquired pneumonia (CAP) in Emergency Departments (EDs). The low-intensity implementation arm reflected the quality improvement methods typically used by the collaborating state quality improvement (QI) organizations, which served as a usual care control. The moderate intensity arm also conducted an on-site educational session for ED medical providers and requested a QI plan for the admission decision. The high-intensity intervention added a multifaceted set of provider behavior change techniques (i.e. reminder forms, feedback, and bimonthly plan-do-study-act cycles) that continued through the year-long guideline implementation period. The intervention was randomized at the site (ED) level in the ratio of 3 high intensity: 3 moderate intensity: 2 low intensity.

The study guideline recommended outpatient care for low risk patients with CAP who presented for the ED and inpatient care for high risk patients. The study practice guideline

recommended four processes of care for outpatients (i.e., oxygen assessment, first dose of antibiotics in ED, treatment with compliant antibiotics in the ED, and compliant antibiotic therapy upon discharge) and four processes of care for inpatients (oxygen assessment, blood cultures before antibiotic administration, antibiotic administration within 4 hours and treatment with compliant antibiotic therapy in ED). Those eight binary outcomes were used to illustrate the proposed method, with "1" indicating that the patient received the recommended process of care, "0" otherwise.

The EDCAP study has a 3-level hierarchical data structure, in which multiple outcomes are nested within patients, patients are nested within providers, and providers are nested within clinical sites (EDs). Of the 3201 patients were seen by 407 providers at 32 clinical sites, 1125 received outpatient care and 2076 received inpatient care. Intervention was randomized at site level.

The primary study results were published by analyzing each outcome separately (Yealy, Auble, Stone et al., 2005). Many pairwise comparisons were conducted and subjective conclusions of the overall effect were based on the multiple tests. The information derived from these individual outcome analyses can be overwhelming, and the increased Type I error is an issue. Further, in this multidimensional outcome data, we expect some relationships between the outcomes, so that modeling outcomes independently could result in a loss of efficiency.

The purpose of our work is to develop a statistical model to assess an overall intervention effect. Specifically, we propose a general latent variable model to analyze the intervention effect with regard to multiple binary outcomes with a 3-level hierarchical data structure, using a fully Bayesian approach. This proposed model framework incorporates the correlation structure into one or more latent outcomes, and simultaneously regresses the latent outcome(s) on

interventions. This analysis will address the question of whether there is significant variability in the quality of care across sites, and whether the intervention explains some of this variability. This latent modeling approach provides a comprehensive way to analyze intervention effect in multivariate and hierarchical data.

In Chapter 2, we review relevant literature. In Chapter 3, we describe a Bayesian formulation of the proposed one-latent trait and two-latent trait models. The EDCAP data is analyzed sequentially in terms of complexity. First, each outcome was analyzed individually using both a Bayesian approach and maximum likelihood with random effects to account for clustering effects at the site and provider levels, and results were presented in Appendix A. Second, in Chapter 4, a one-latent trait model is proposed to assess an overall intervention effect using outpatient data. Then, in Chapter 5, the one latent trait is extended to two-latent trait model to assess the intervention effect on both outpatient care (4 outcomes) and inpatient care (4 outcomes). We conclude in Chapter 6 by discussing some limitations and extensions. Selected output and figures of Bayesian implementation are listed in Appendix B.

2.0 REVIEW OF THE LITERATURE

Multiple outcomes are common in the social and behavior sciences, and increasingly, so in biomedical studies. Statistical analysis is challenged by such data in studying the relationships between outcome measures and associated covariates. The development of statistical methods for the analysis of multiple outcomes has been an area of active research for decades.

2.1 NON LATENT VARIABLE APPROACHES

Traditionally, one frequently used approach is to analyze the treatment effect on each response variable separately, presenting multiple P-values and drawing overall subjective conclusions. While this approach is simple and easy to implement, it has limitations. Evaluating individual response variables is informative but often fails to provide an overall statement of the treatment effect. It is commonplace to interpret a trial as positive if any endpoint has a treatment difference significant at the 5% level, which increased the risk of overall Type I error rate (inflation of false positives). Also, it fails to advantageously borrow strength across the response variables, i.e., it fails to combine the related information about the treatment effect from the various response variables. Hence, for multiple endpoints without prespecified priorities, it is challenging to preserve a small overall Type I error rate and allow for correlated endpoints. One could simply apply a Bonferroni correction; however, this approach can be very conservative, particularly for

correlated endpoints. For related outcomes likely to be affected in a similar manner, a multivariate approach that allows for an overall assessment of intervention effects by combining information from these related outcomes is preferred, which could detect the intervention effects with better statistical power (Gray SM and Brookmeyer R 1998, Sammel M, Lin H and Ryan L 1999)

Another commonly used approach is to reduce the dimensionality of the data. Very simple approaches include calculating a summation score for several continuous variables or collapsing multiple binary outcomes into a single binary outcome that indicates the presence of at least one of the endpoints of interest. This could also been accomplished in a more complex way using variable reduction techniques, such as factor analysis or principal components (Morrison, 1976). This analysis apporach is named as two-stage factor analysis, a two-step procedure, wherein one first performs a factor analysis or principal components analysis to identify a linear combination of outcomes that are most correlated with each other, then treats this linear combination as fixed and known without measurement error and models this linear combination as a function of covariates (Sammel, 1999). This approach may not lead to interpretable results and the data reduction may result in biased estimates, a loss of information (Gray SM and Brookmeyer R 1998, Croon Bolck, 1997), and most important, uncertainty in the aggregated scores is difficult to quantify (Gray and Brookmeyer, 1998).

Another approach is the use of global testing procedures based on suitable multivariate models to compare outcomes between different groups. Testing procedures for multiple outcomes have been described by O'Brien (1984), Pocock (1987) and Legler (1995) among others. O'Brien (1984) based his methods on generalized least squares (GLS) as well as a nonparametric rank-based approach. Pocock et al. (1987) extended O'Brien's GLS approach to

included binary and survival endpoints. Legler et al. (1995) used generalized estimating equation (GEE)-based score tests for a general M-group comparison. Although a global test can assess whether or not groups differ with respect to multiple outcomes, they fall short in exploring the relationship between outcomes. There is no well-recognized best way to accomplish this. One simple approach is to use some kind of multiple comparisons technique to identify which individual outcomes are affected. Ironically, attempting to do so may sometimes undermine the rationale for using a global test in the first place. Global analyses have been criticized often on the grounds that combining multiple outcomes may obscure a real effect if only one or two outcomes are actually affected by the intervention. In general, global tests will be more powerful and are recommended in practice when all the outcomes are closely related and be similarly affected by the variable of interest.

None of these approaches described above provide a comprehensive way to study the relationships between outcomes, to synthesize (or summarize) these variables, or to quantify the multivariate outcome as a univariate composite score and simultaneously assess the relationship between this composite score and the observed covariates with measurement errors considered in one model framework. However, latent variable models provide a natural and comprehensive way to analyze data with multiple outcomes and hierarchical data structure.

2.2 LATENT VARIABLE MODELS

Latent variable models refer to any models that include unobserved random variables. Such models assume the existence of one or more latent variables, i.e., quantities that are not directly observed but thought to be underlying the measured responses. Latent variable models have

been used extensively in the areas educational testing, psychology, and social science for studying interrelationships among observed and latent variables, because they provide a natural way to analyze data with multiple dimensions. The utility of these models in medical research has only quite recently been recognized (Bentler PM and Stein JA 1992, Rabe-Hesketh S. and Skrondal A. 2008). One advantage of the latent variable model is that a one degree-of-freedom test can be used to test for the overall exposure effect, and could be more powerful compared to the a M degree-of-freedom test (Sammel 1999). Another advantage of the latent variable model is that it naturally yields a summary measure for each individual that can be interpreted as individual severity score.

The classification scheme of traditional latent variable models was summarized by Skrondal A. and Rabe-Hesketh S (2007), based on metrics for the observed and latent variables. For both continuous latent variable(s) and observed variables, the models include common factor models, structural equation models, linear mixed models and covariate measurement error models; latent trait models/IRT for continuous latent variable(s) and categorical observed variables; latent profile models for categorical latent variable(s) and continuous observed variables; and latent class models for both categorical latent variable(s) and observed variables

The factor analysis model is the most basic statistical model for studying the relationships among latent and observed variables. To deal with complex data sets in various fields, the factor analysis model has been generalized to more sophisticated models (Bentler and Weeks 1980). These multivariate models are commonly called structural equation models (SEMs). A SEM with latent variables provides a very general framework for modeling the relationships in multivariate data (Bollen 1989). In the following, in order to have a better understanding of the proposed latent variable models, we will review various latent variable

models, including factor models, item response models, SEMs, and the recent work on unifying and extending the classical latent variable models within a general framework. We also review Bayesian estimation for latent variable models.

2.2.1 Common factor models

For $i = 1, \ldots, N$ independent subjects with $j=1,\ldots,$ I continuous items, a unidimensional common factor model can be written as

$$y_{ij} = \beta_j + \lambda_j \eta_i + \varepsilon_{ij}, \text{ where } \eta_i \sim N(0,\psi), \varepsilon_{ij} \sim N(0,\theta_{jj}) \, , Cov(\eta_i, \varepsilon_{ij}) = 0 \qquad (2.1)$$

In (2.1), β_j represents the measure-specific mean, η_i is the common factor or latent trait for subject i, λ_j is a factor loading for the j^{th} item, ε_{ij} are unique factors or measurement error (which are not separately identified when there are no replicates) and θ_{jj} is the measurement error variance. The model framework could be considered as a multivariate model that incorporates the correlation structure into a single latent outcome, which is a weighted combination of the observables. The common factor can represent any hypothetical construct, i.e., a concept that cannot even in principle be directly observed, intelligence and depression being prominent examples. In this case the measures j are typically questions or items of a questionnaire or structured interview. The answer to a particular item is therefore a reflection of both the hypothetical construct and an item-specific aspect, referred to as the common and specific factors, respectively.

Without any parameter constraints, the above unidimensional common factor model is not identified (several sets of parameter values can produce the same probability distribution) because multiplying the standard deviation $\sqrt{\psi}$ of the common factor by an arbitrary positive

constant can be counteracted by dividing all factor loadings λ_j by the same constant. Identification is achieved either by 'anchoring', where the first factor loading is fixed to one ($\lambda_1 = 1$), or by 'factor standardization', where the factor variance is set to a positive constant ($\psi = 1$). The models resulting from either identification restriction are equivalent.

Hypothetical constructs are often multidimensional. The above one-dimensional factor model could be generalized to a less restrictive multi-dimensional ("M-dimensional") factor model. An M-dimensional factor model can be formulated as

$$y_{ij} = \beta_j + \lambda_{j1}\eta_{i1} + ... + \lambda_{jM}\eta_{iM} + \varepsilon_{ij} \qquad (2.2)$$

In matrix form $y_i = \beta + \Lambda\eta_i + \varepsilon_i$

where β is a vector of item-specific constants, Λ is now an $n \times m$ matrix of factor loadings with element pertaining to item j and latent variable l denoted λ_{jl}, η_i is a vector of M common factors with covariance matrix Ψ, and ε_j is a vector of unique factors with diagonal covariance matrix Θ.We define $\Psi \equiv \text{Cov}(\boldsymbol{\eta}_i)$ and assume that $\text{E}(\boldsymbol{\eta}_j) = \mathbf{0}$, $\text{E}(\varepsilon_i) = \mathbf{0}$, and $\text{Cov}(\boldsymbol{\eta}i, \varepsilon_i) = \mathbf{0}$.

In the multidimensional case, an important example of a restricted model is the *independent clusters* model where Λ has many elements set to zero such that each indicator measures one and only one factor. Such a configuration makes sense if one set of indicators is designed to measure one factor and another set of indicators to measure another factor. For example, matrix (2.3) in below depicts an independent clusters two-factor model where each factor is measured by three separate items. The model identification is achieved by anchoring, where we have the scale of each factor by setting one factor loading to 1.

$$\begin{bmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \\ y_{i4} \\ y_{i5} \\ y_{i6} \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ 0 & 1 \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix} \begin{bmatrix} \eta_{i1} \\ \eta_{i2} \end{bmatrix} + \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \\ \varepsilon_{i5} \\ \varepsilon_{i6} \end{bmatrix} \tag{2.3}$$

2.2.2 IRT models

When the response items are dichotomous or ordinal, the above factor model can be extended to an IRT model using a generalized linear model formulation. The conditional probability of a particular response given the latent trait (or factor) typically is specified by a logit or probit link function. The IRT model has been developed in the context of educational testing. Let y_{ij} denotes the observed binary outcome of item j (correct or incorrect) on subject *i*, and η_i represents the continuous unobserved ability of the examinee. A two-parameter logistic (2-PL) model can be formulated as

$$\Pr(y_{ij} = 1 \mid \eta_i) = \frac{\exp(\lambda_j \eta_i - \beta_j)}{1 + \exp(\lambda_j \eta_i - \beta_j)} \tag{2.4}$$

There are two parameters associated with each item, an intercept and a factor loading. β_j represents the *item difficulty,* η_i represents ability, λ_j is referred to as the discrimination parameter because items with a larger λ_j better discriminates between subjects with different abilities. In this 2-PL model, an item can be easier than another item for low abilities but more difficult than the other item for higher abilities, due to the item-examinee interaction $\lambda_j \eta_i$. The two-parameter (2-PL) IRT model could be reduced to a one-parameter (1-PL) IRT model by

constraining factor loadings λ_j to all be 1, which is just a random intercept logistic model without covariates. The Rasch model could be obtained by taking η_i as fixed in the 1-PL model.

The two-parameter IRT model with probit link is called the Normal-ogive model (Lord FM 1952)

$$\Phi^{-1}[\Pr(y_{ij}=1|\eta_j)] = \beta_i + \lambda_i\eta_j \tag{2.5}$$

Here $\Phi(\cdot)$ is the cumulative standard normal distribution and $\Phi^{-1}(\cdot)$ is the probit link.

2.2.3 Structural equation models (SEM)

SEMs, also called simultaneous equation models, are multivariate regression models with latent variables. SEMs provide a broad framework for modeling means and covariance relationships in multivariate data. In general, SEM combines the ideas of factor analysis and regression. Unlike the more traditional multivariate linear model, however, the response variable in one regression equation in a SEM may appear as a predictor in another equation. Factor models and IRT models are important in their own right for modeling the relationship among the observed and latent variables, but also as building blocks in SEMs, where relationships among latent variables are modeled. SEM also is referred to as covariance structure analysis. In this structural model, there could be both latent dependent variables and latent explanatory variables.

The standard SEM, in particular the LISREL (linear structural relationship) model (Jöreskog KG 1977), is composed of two components. The first component is a confirmatory factor analysis model, which relates the latent variables to their entire corresponding manifest (observed) variables (indicators) and takes the measurement error into account. This component can be regarded as a regression model that regresses the manifest variables on a small number of

latent variables. The second component is again a regression type structural equationthat regresses the endogenous (dependent) latent variables with the linear terms of some endogenous and exogenous (independent) latent variables. There are several ways of parameterizing SEMs with latent variables. Here we will use a parameterization suggested by Muthén (Muthén BO 1984) due to its convenience in application.

The measurement part of the model is the confirmatory factor model specified in equation (2.1)

$$y_{ij} = \beta_j + \lambda_{j1}\eta_{i1} + ... + \lambda_{jM}\eta_{iM} + \varepsilon_{ij} \tag{2.6}$$

The structural part of the model specifies regressions for the latent variables on other latent and observed variables

$$\eta_i = \alpha + \mathrm{B}\eta_i + \Gamma\chi_i + \zeta_i \tag{2.7}$$

Here η_i is a vector of latent variables with corresponding lower-triangular parameter matrix **B** governing the relationships among them, α is a vector of intercepts, Γ a regression parameter matrix for the regression of the latent variables on the vector of observed covariates χ_i, and ζ_i is a vector of disturbances. We define $\Psi \equiv$ Cov$(\boldsymbol{\zeta_i})$ and assume that E$(\boldsymbol{\zeta_i})$=**0**, Cov$(\chi_i, \boldsymbol{\zeta_i})$ = **0** and Cov$(\varepsilon_i, \boldsymbol{\zeta_i})$ = **0**,

An important special case is the Multiple-Indicator-Multiple-Cause (MIMIC) model (Joreskog KG 1975), which imposes the restriction **B = I** in the structural model, so the structural model reduces to

$$\eta_i = \alpha + \Gamma x_i + \zeta_i \tag{2.8}$$

The MIMIC model is a one-factor model, where the factor is measured by multiple indicators and regressed on several observed covariates or “causes”.

2.2.4 Generalized Linear Latent and Mixed Models

Recognizing the mathematical similarity of a wide range of latent variable models, Rabe-Hesketh and Skrondal proposed a very general framework to unify and generalize various the classical latent variable models, including the multilevel, factor, item response, latent class, structural equation and longitudinal models. The model framework is named as Generalized Linear Latent and Mixed Models, or GLLAMMs (Rabe-Hesketh et al., 2007; Skrondal & Rabe-Hesketh, 2004). The GLLAMM model can be written down explicitly in its full generality just like the unifying LISREL model.

GLAMM combines features of generalized linear mixed models (GLMMs) and SEMs. GLLAMMs consist of two building blocks: a response model and a structural model. The response model specifies the distribution of the observed responses conditional on the latent variables and covariates (via a linear predictor and link function) and in the structural model the latent variables themselves maybe regressed on other latent and observed covariates.

The response model generalizes GLMMs to incorporate factor structure in addition to random intercepts and coefficients. Conditional on the latent variables, the response model of many latent variable models is a generalized linear model. As for such models, the response model of GLLAMMs has three components: a link, a distribution and a linear predictor. The conditional expectation of the response y, given $\mathbf{x}$, $\mathbf{z}$ and η, is 'linked' to the linear predictor via a link function $g(\cdot)$.

$$g(E[y_{ij} \mid x, z, \eta]) = X_{ij}`\beta + \sum_{m=1}^{M} \eta_{mi} z`_{mij} \lambda_m \tag{2.9}$$

Common combinations of links and distributions include: (i) the identity link and normal distribution for continuous responses; (ii) the logit, probit or complementary log–log link and

Bernoulli distribution for dichotomous responses; (iii) the cumulative version of these links and multinomial distribution for ordinal responses; and (iv) the log link and Poisson distribution for counts. The right side of the equation is the linear predictor, η is the vector of all latent variables in the model, **x** and **z** denote vectors of covariates. The elements of $\mathbf{x}_{ij}$ are covariates with 'fixed' effects β. The m^{th} latent variable η_{mj} is multiplied by a linear combination $z`_{mij}\lambda_m$ of covariates $z`_{mij}$ where λ_m are parameters (usually factor loadings). Some traditional latent variable models could be viewed as special cases, such as the common factor model and IRT models.

The structural model is similar to the structural part of SEM specified in (2.7) except that it is a *multilevel* structural model where latent variables and observed variables can vary at different hierarchical levels. This includes the conventional single-level structural model as a special case.

2.2.5 Bayesian estimation of latent variable models

Most of the literature on latent variable models is frequentist in nature, based on maximum likelihood estimation (MLE) (for examples see Bock and Aitkin, 1981, Rabe-Hesketh et al., 2004 and 2005, Schoenberg and Richtand, 1984). Latent variables are treated as random and parameters are fixed, inference is usually based on the marginal likelihood, the likelihood of the data given the latent variables, integrated (or summed in the discrete case) over the latent variables distribution. In the cases of multivariate normal latent variables and multivariate normal responses, the marginal distribution of the response given the latent variables is multivariate normal (multivariate normal marginal distribution), the MLE is relatively straightforward because the integral involved in the marginal likelihood can be explicitly solved

and expressed in closed form. However, in non-normal models, such as item-response models, there generally is no analytic expression for the likelihood, i.e., the likelihood does not have a closed form, and approximations are needed (Skrondal A and Rabe-hesketh S, 2004). There are several more or less accurate approximate methods of integration, including numerical integration using quadrature or adaptive quadrature (as implemented in SAS NLMIXED and the Stata program gllamm), or Monte Carlo integration (simulated likelihood). Different methods for maximizing the likelihoods include the Expectation-Maximization (EM) and Gradient methods of Newton-Raphson and Fisher Scoring algorithms.

In contrast, Bayesian estimation treats both latent variables and parameters as random variables, so that there is no distinction between the two types of variables. When the likelihood does not have a closed form, the Bayesian alternatives using MCMC for latent variable models are attractive (Sik-Yum Lee 2007a, Sik-Yum Lee 2007b). MCMC methods allow estimation of a very wide range of models and have become increasing popular. The difficulties induced by the complexities of the model and the multi-dimensional integrals can be handled efficiently by means of powerful computing tools in statistics, such as the Gibbs sampler (Geman and Geman, 1984). An additional benefit is that samples are available from the joint posterior distribution of the latent variables.

A major breakthrough for Bayesian estimation of latent variable models is the idea of data augmentation that was proposed by Tanner and Wong (1987). The strategy is to treat latent quantities (Ω) as hypothetical missing data and augment the observed data with them so that the posterior distribution based on the "complete" data is relatively easy to analyze. More specifically, instead of working with the intractable posterior density $P(\theta \mid y)$, we will work with $P(\theta, \Omega \mid y)$, where Ω is the set of latent variables in the model. For most cases, $P(\theta, \Omega \mid y)$

still is not in closed form and it is difficult to deal with it directly. However, on the basis of the complete-data set (Ω, y), the conditional distribution $[\theta \mid \Omega, y]$ is usually standard; moreover, the conditional distribution $[\Omega \mid \theta, y]$can be derived from the definition of the model without much difficulty. Consequently, we can apply some MCMC methods to simulate the observations from $P(\theta, \Omega \mid y)$ by drawing observations iteratively from their full conditional densities $p(\theta \mid \Omega, y)$ and $p(\Omega \mid \theta, y)$ (Sik-Yum Lee 2007a). Most of the full conditional distributions are the standard normal, gamma or Wishart distributions. Simulating observations from them is rather straightforward. For nonstandard conditional distributions, the Metropolis-Hastings (MH) algorithm (Metropolis el al., 1953; Hastings, 1970) may have to be used for efficient simulation.

The freely available software WINBUGS (Windows version of Bayesian inference Using Gibbs Sampling, Gilks WR 1994) is a useful and powerful tool for Bayesian analysis. It allows users to evaluate complex models without implementing the technical details of MCMC for each new problem. The software will set up the MCMC process and conduct the sampling, so that the users are able to focus on model design and validation. It could produce reliable Bayesian statistics for a wide range of statistical models, including SEM models (Sik-Yum Lee 2007a,). The algorithm used in WinBUGS is mainly developed using MCMC techniques, such as the Gibbs sampler and the MH algorithm. This software is able to produce reliable Bayesian statistics, including the Bayesian estimates with their standard error estimates, and the Deviance Information Criterion for model comparison and goodness-of-fit assessment of the hypothesized model.

3.0 PROPOSED METHOD

We describe our method in the context of EDCAP example. The EDCAP study has a 4-level hierarchical data structure, in which multiple measurements are nested in patients, patients are nested in providers and providers are nested in clinical sites. The intervention is randomized at the site level. One question of interest is whether the intervention is association with site level variation. Assuming that each site has one underlying trait (e.g., quality of pneumonia care), we construct a latent trait at site-level and then model the association between this latent trait and the intervention.

To set the notation, let *i* denote the level-4 units (sites), *j* denote the level-3 units (providers), *k* denote the level-2 units (patients), and *h* denote the level-1 units (measurements or items). Let Y_{ijkh} be the dichotomous value of the h^{th} response of the k^{th} patient treated by provider *j* in the i^{th} (i=1,2,…., 32) site (ED). The two covariates that will be modeled are one level-2 unit (patient-level) covariate, risk status (Z_{ijk} 0=low risk, 1=high risk)), and one level-4 unit (site-level) covariate, intervention level ($X_i(x_{i1}, x_{i2})$ *(0,0)*=low, (1,0)=moderate, (1,1)=high intensity intervention).

3.1 ONE-LATENT TRAIT MODEL

EDCAP outpatient data with four outcomes were used to illustrate the proposed method

3.1.1 Model specification

Though data are collected at patient level, this analysis focuses on site-level variation. We formalize the likelihood function by writing the within-site and the between-site models separately, with the within-site model linking the outcomes to patient risk status and various random effects, and between-site model linking the site-level random effect to the latent variable, with the latent mean modeled as a function of the intervention.

Within-Site Model. The outcome variable is assumed to follow a binomial distribution and is linked by logit link to the patient-level covariates, a patient-level random effect, a provider-level random effect, and a site-level measurement-specific random effect,

$$Logit(y_{ijkh} \mid \beta_{0h}, \beta_{1h}, \mu_{ijk}, \upsilon_{ij}, \eta_{ih}, z_{ijk}) = \beta_{0h} + \beta_{1h} z_{ijk} + \mu_{ijk} + \upsilon_{ij} + \eta_{ih} \qquad (3.1)$$

where,

Y_{ijkh}, binary effectiveness outcome measure

β_{0h} , a baseline for each measure, indicating performance of the h^{th} process of care at an average site with the low-intensity intervention for low risk patients. The lower the value, the lower the average probability of receiving the h^{th} process of care.

$\beta_{1h,}$ a fixed measure-specific regression coefficient for the patient-level covariate

$Z_{ijk,}$ as defined above, patient risk status (i.e., 0= high risk, 1= low risk)

The parameter μ_{ijk} is patient level random error term such that the responses are conditionally independent at the patient level. It is assumed to be normally distributed with zero mean and constant variance (i.e., $\mu_{ijk} \sim N(0, \sigma^2)$)

The parameter ν_{ij} is a provider level random error term that allows for correlation among patients within a provider. It is assumed to be normally distributed with zero mean and constant variance (i.e., $\nu_{ij} \sim N(0, \tau^2)$.

The parameter η_{ih}, a measurement-specific site level random effect for measure h in site i, permits heterogeneity across the sites and allows for correlation among patients within a site. It is assumed to be normally distributed with a non-zero mean as a function of intervention effect and a measurement-specific variance (i.e., $\eta_{ih} \sim N((\lambda_h\theta_i, \psi_h^2))$, which are specified in detail below. It is further assumed that the random effects μ_{ijk} , ν_{ij} and η_{ih} are mutually independent.

Between-Site Model. To model the correlation between the multivariate outcomes at the site-level, we relate the site-level random effects to one latent variable θ_i, such that the four site-level effects are conditionally independent given the latent variables.

$$\begin{aligned} &\eta_{ih} = \lambda_h\theta_i + \varepsilon_{ih}, \text{ where } \varepsilon_{ih} \sim N(0, \psi_h^2) \\ &\textit{i.e. } \eta_{ih} \sim N(\lambda_h\theta_i, \psi_h^2) \end{aligned} \tag{3.2}$$

Where, θ_i serves as a composite profile of care rendered by site i. Larger θ_i corresponds to better quality of health care.

The parameter λ_h is a fixed measure-specific discrimination parameter (weight), depending on a measure's ability to discriminate between sites. It may be thought of as a "factor loading" that quantifies the weight of each outcome on the latent variables. The larger the value, the more the power to discriminate underlying quality. The sign of λ_h is not identifiable, so the constraint that $\lambda_h > 0$ is added to the model.

The variable $\varepsilon_{ih} \sim (0,\psi_h^2)$, the site level random error term for measure *h*, measures the degree of heterogeneity across sites.

In this hierarchical model, variability in latent quality (θ_i) across sites is assumed to have both a systematic component, explained by a site-specific covariate (the intervention X_i: low, moderate, high intensity), and a random component such that

$$\begin{aligned} &\theta_i = \gamma X_i + \alpha_i \text{ where } \alpha_i \sim N(0,1) \\ &i.e.\ \theta_i \sim N(\gamma X_i, 1) \end{aligned} \tag{3.3}$$

where variable α_i is *i.i.d* error term (i.e. $\alpha_i \sim$ N (0,1)), so that the prior variance of the latent quality trait is set to 1 (to fix the scale of the latent variable for identifiability of the model). Estimates of the latent trait ($\theta_{i.}$) are given by the posterior mean of θ_i. Low values indicate poor quality of care.

Parameter γ is the vector of fixed treatment effects, $\gamma = (\gamma_1 , \gamma_2)$ with γ_1 denoting moderate intensity vs. low intensity, and γ_2 denoting high intensity vs. low intensity. The linear contrast $\gamma_1 - \gamma_2$ denotes high intensity vs. moderate intensity.

Here X_i denotes the vector of site-level covariates, $X_i = (x_{i1}, x_{i2})$, where (0, 0) = low intensity intervention, (1,0) = moderate intensity intervention, (1,1) = high intensity intervention.

It follows from the two-stage between-site models (specified above in equations 3.2 and 3.3) that we can formalize the between-site model in one equation

$$\begin{aligned} &\eta_{ih} = \lambda_h \gamma X_i + \lambda_h \alpha_i + \varepsilon_{ih}, \text{ where } \alpha_i \sim N(0,1), \varepsilon_{ih} \sim N(0, \psi_h^2) \\ &i.e.\ \eta_{ih} \sim N(\gamma X_i, \lambda_h^2 + \psi_h^2) \end{aligned} \tag{3.4}$$

Based on the two submodels specified above in equation 3.1 and 3.4, the full model can be written as

$$Logit(y_{ijkh} \mid \beta_{0h}, \beta_{1h}, \lambda_h, \theta_i, \mu_{ijk}, \upsilon_{ij}, \varepsilon_{ih}, z_{ijk}) = \beta_{0h} + \beta_{1h} z_{ijk} + \mu_{ijk} + \upsilon_{ij} + \lambda_h \theta_i + \varepsilon_{ih} \tag{3.5}$$

In matrix form

$$\begin{bmatrix} Logit(y_{ijk1}) \\ Logit(y_{ijk2}) \\ Logit(y_{ijk3}) \\ Logit(y_{ijk4}) \end{bmatrix} = \begin{bmatrix} \beta_{01} \\ \beta_{02} \\ \beta_{03} \\ \beta_{04} \end{bmatrix} + \begin{bmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{13} \\ \beta_{14} \end{bmatrix} Z_{ijk} + \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix} [\theta_i] + \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \end{bmatrix} + \nu_{ij} + \mu_{ijk}$$

where $\theta_i = \gamma X_i + a_i$ as specified in equation 3.3

$$\mu_{ijk} \sim N(0, \sigma^2) , V_{ij} \sim N(0, \tau^2), \varepsilon_{ih} \sim N(0, \psi_h^{2)}$$

or formalized in one equation

$$Logit(y_{ijkh} \mid \beta_{0h}, \beta_{1h}, \lambda_h, \gamma, \mu_{ijk}, \upsilon_{ij}, \varepsilon_{ih}, z_{ijk}, X_i) = \beta_{0h} + \beta_{1h} z_{ijk} + \mu_{ijk} + \upsilon_{ij} + \lambda_h \gamma X_i + \lambda_h \alpha_i + \varepsilon_{ih} \quad (3.6)$$

where $\mu_{ijk} \sim N(0, \sigma^2), \upsilon_{ij} \sim N(0, \tau^2), \alpha_i \sim N(0,1)$, and $\varepsilon_{ih} \sim N(0, \psi_h^2)$

There are a total of 20 unknown parameters in this model for 4 outpatient measures, including treatment effect $\gamma(\gamma_1, \gamma_2)$, measurement specific intercept (β_{0h}), measurement- specific factor loading (λ_h), measurement- specific covariate effect (β_{1h}) , and the variance σ^2 of the patient -level random error term(μ_{ijk}), the variance $\tau^2$of the provider-level random error term(v_{ij}) and the variance ψ_h^2 of the measurement- specific site-level random error terms ($\underline{\varepsilon_{ih}}$). A path diagram of this model is given in Figure 3-1.

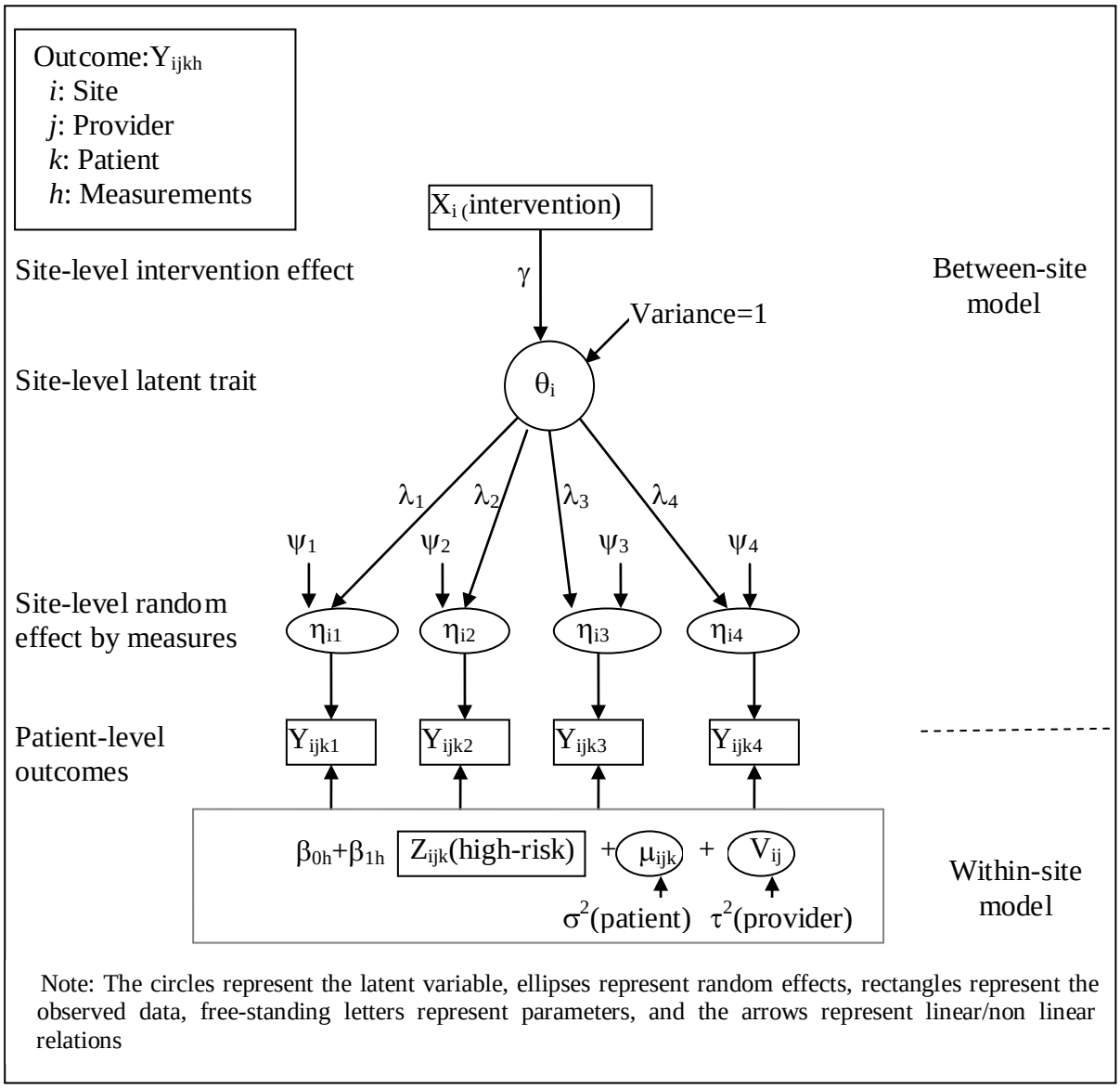


Figure 3-1 Path diagram of one latent trait model with four outcomes

3.1.2 Model estimation

We used a fully parametric Bayesian approach for model estimation. Bayesian estimation of the model parameters requires the specification of a prior distribution for each unknown parameter. In this study, a noninformative but proper prior distribution is used. We assume that the fixed effects ($\{\beta_{0h}\}$, $\{\beta_{1h}\}$ and $\{\gamma\}$) are independent and normally distributed with mean zero and a huge variance 10^4 (N (0, 10^4). The prior distributions for factor loadings $\{\lambda_h\}$ also were chosen to be N (0, 10^4), but truncated below 0 for model identification. For the variance parameters, we follow the recommendation of Andrew Gelman (2006) to use a noninformative uniform prior density on standard deviation parameters unless a weakly informative prior is desired. The uniform density on σ is equivalent to $p(\sigma^2) \propto \sigma^{-1}$, an inverse χ^2 distribution with -1 degree of freedom. We used Uniform (0,100) as prior for the standard deviation of σ, τ, and $\{\psi_h\}$ to account for clustering at the patient, provider and site levels. The inverse-gamma (ε,ε) family of noninformative prior distributions is not recommended here, because when σ is estimated to be near zero the resulting inference will be sensitive to ε. The setting of near-zero variance parameters is important, partly because this is where the classical and Bayesian inference for hierarchical models will differ the most. To identify parameters, we fix the scale of the latent variable by setting the variance of the prior distribution for θ_i equal to 1.

Bayesian estimates of latent variable models can be obtained using MCMC techniques, including Gibbs sampler (Geman and Geman, 1984) and the Metropolis Hasting (MH) algorithm (Metropolis et al., 1953; Hasting, 1970). The strategy is to treat latent quantities (Ω), including the latent trait and random effects, as hypothetical missing data; due to the nature of MCMC, it is not necessary to integrate out the latent quantities to make inference about the parameters. The

latent quantities are updated, along with other parameters from their posterior distributions $P(\theta,\Omega \mid y)$ by drawing observations iteratively from their full conditional densities $p(\theta \mid \Omega, y)$ and $p(\Omega \mid \theta, y)$. For the proposed model specified by equations (3.1-3.6), the augmented complete data likelihood function takes the form:

$$
\begin{aligned}
&f(Y,\Omega \mid \beta_0,\beta_1,\lambda,\gamma,\sigma,\tau,\psi,Z,X) = \prod_{i=1}^{32}\prod_{j=1}^{n_{ij}}\prod_{k=1}^{n_{ijk}}\prod_{h=1}^{4} f(y_{ijkh},\theta_i,\mu_{ijk},\upsilon_{ij},\varepsilon_{ih} \mid \beta_{0\mathrm{h}},\beta_{1\mathrm{h}},\lambda,\sigma,\tau,\psi_h,z_{ijk},X_i) \\
&= \prod_{i=1}^{32}\left(\prod_{j=1}^{n_{ij}}\left(\prod_{k=1}^{n_{ijk}}\left(\prod_{h=1}^{4} f(y_{ijkh} \mid \beta_{0\mathrm{h}},\beta_{1\mathrm{h}},\lambda,\theta_i,\mu_{ijk},\upsilon_{ij},\varepsilon_{ih})\right)\phi(\mu_{\mathrm{ijk}};0,\sigma)\right)\phi(\upsilon_{\mathrm{ij}};0,\tau)\right)\phi(\theta_i;X_i'r,1)\prod_{h=1}^{4}\phi(\varepsilon_{\mathrm{ih}};0,\psi_h) \\
&\propto \prod_{i=1}^{32}\prod_{j=1}^{n_{ij}}\prod_{k=1}^{n_{ijk}}\prod_{h=1}^{4}\frac{\exp\{y_{ijkh}(\beta_{0\mathrm{h}}+\beta_{1\mathrm{h}}z_{\mathrm{ijk}}+\mu_{\mathrm{ijk}}+\upsilon_{\mathrm{ij}}+\lambda_{\mathrm{h}}\theta_{\mathrm{i}}+\varepsilon_{\mathrm{ih}})\}}{1+\exp\{y_{ijkh}(\beta_{0\mathrm{h}}+\beta_{1\mathrm{h}}z_{\mathrm{ijk}}+\mu_{\mathrm{ijk}}+\upsilon_{\mathrm{ij}}+\lambda_{\mathrm{h}}\theta_{\mathrm{i}}+\varepsilon_{\mathrm{ih}})\}} \qquad (3.7) \\
&\times \prod_{i=1}^{32}\prod_{j=1}^{n_{ij}}\prod_{k=1}^{n_{ijk}}\left(\frac{1}{\sigma}\right)^{\frac{1}{2}}\exp\{\frac{1}{2\sigma}(\mu_{\mathrm{ijk}})^2\} \times \prod_{i=1}^{32}\prod_{j=1}^{n_{ij}}\left(\frac{1}{\tau}\right)^{\frac{1}{2}}\exp\{\frac{1}{2\tau}(\nu_{\mathrm{ij}})^2\} \\
&\times \prod_{i=1}^{32}\exp\{-\frac{1}{2}(\theta_i-\gamma X_i)^2\}\times\prod_{i=1}^{32}\prod_{h=1}^{4}\left(\frac{1}{\psi_h}\right)^{\frac{1}{2}}\exp\{\frac{1}{2\psi_h}(\varepsilon_{ih})^2\}
\end{aligned}
$$

Given the augmented complete data likelihood in (3.7) and priors specified as above, the joint posterior of latent quantities and all unknown parameters can be expressed as

$$
\begin{aligned}
&f(\beta_0,\beta_1,\lambda,\gamma,\sigma,\tau,\psi,\theta_i,\mu_{ijk},\upsilon_{ij},\varepsilon_{ih} \mid Y,X,Z) \qquad (3.8) \\
&\propto \{\prod f(y_{ijkh},\theta_i,\mu_{ijk},\upsilon_{ij},\varepsilon_{ih} \mid \beta_{0\mathrm{h}},\beta_{1\mathrm{h}},\lambda,\sigma,\tau,\psi_h)\} f(\beta_0,\beta_1,\lambda,\gamma,\sigma,\tau,\psi) \\
&= \{\prod_{i=1}^{32}\left(\prod_{j=1}^{n_{ij}}\left(\prod_{k=1}^{n_{ijk}}\left(\prod_{h=1}^{4} f(y_{ijkh} \mid \beta_{0\mathrm{h}},\beta_{1\mathrm{h}},\lambda,\theta_i,\mu_{ijk},\upsilon_{ij},\varepsilon_{ih})\right)\phi(\mu_{\mathrm{ijk}};0,\sigma)\right)\phi(\upsilon_{\mathrm{ij}};0,\tau)\right)\phi(\theta_i;X_i'r,1)\prod_{h=1}^{4}\phi(\varepsilon_{\mathrm{ih}};0,\psi_h)\ \} \\
&\quad \times f(\beta_0,\beta_1,\lambda,\gamma,\sigma,\tau,\psi)
\end{aligned}
$$

The parameters are sampled using the augmented data. The full conditional distributions needed to implement the MCMC algorithms are summarized below:

(1) Sample θ_i using the Metropolis algorithm from the full conditional distribution:

$$Pr(\theta_i|.\ \propto) Pr(y_{ijk}|\beta_0,\beta_1,\lambda,\sigma,\tau,\psi)p(\theta_i|\gamma)$$

$$\propto \exp\{\sum_{k=1}^{h}[y_{ijk}(\beta_{0h}+\beta_{1h}z_{ijk}+\mu_{ijk}+\upsilon_{ij}+\lambda_h\theta_i+\varepsilon_{ih})\text{-}\log(1+e^{\beta_{0h}+\beta_{1h}z_{ijk}+\mu_{ijk}+\upsilon_{ij}+\lambda_h\theta_i+\varepsilon_{ih}})$$

$$\text{-}\frac{1}{2}(\theta_i - rX_i)^2\}$$

(2) Sample random effects of μ_{ijk}, υ_{ij}, ε_{ih} using rejection sampling (Zeger and Karim 1991)

(3) Sample variance of random effects using inverse gamma distribution

$$f(\sigma^2|.) \propto IG\left((\frac{I+0.01}{2}, \frac{\sum \mu_{ijk}{}^2 + 0.01}{2})\right), \text{ similarly for } \tau^2 \text{ and } \psi_h{}^2$$

(4) Sample $\beta_h(\beta_{0h},\beta_{1h},\gamma_h)$ using the metropolis algorithm from the full conditional distribution:

$$f(\beta_h|.) \propto \prod_{i=1}^{32}\prod_{j=1}^{n_{ij}}\prod_{k=1}^{n_{ijk}}\frac{\exp\{y_{ijkh}(\beta_{0h}+\beta_{1h}z_{ijk}++\upsilon_{ij}+\lambda_h\theta_i+\varepsilon_{ih})\}}{1+\exp\{y_{ijkh}(\beta_{0h}+\beta_{1h}z_{ijk}+\mu_{ijk}+\upsilon_{ij}+\lambda_h\theta_i+\varepsilon_{ih})\}}\exp(-\frac{1}{2}\beta'\sum\beta^{-1}\beta)$$

(5) Sample γ using the full conditional distribution:

$$p(\gamma|.) \propto N((X`X)^{-1}X`\theta),(X`X)^{-1})$$

The Gibbs sampler algorithm proceeds by sampling latent quantities and parameters from (1) to (5), respectively. The simulation of observations from the standard distributions involved in equations for steps 3 and 5 is straightforward. The MH algorithm could be used to simulate observations from the remaining complex distributions. Repeat Steps (1)-(5) until convergence and collect a large number of additional draws from which to calculate posterior summaries.

The generic Bayesian Package WinBUGS1.4 (Spiegelhalter, Thomas, Best and Lunn, 2003) was used to perform MCMC simulations of the posterior distribution. For our model, we used double chains with two sets of varied initial values. In one set we specify 0 as initial values

for the fixed effects ($\{\beta_{0h}\}$, $\{\beta_{1h}\}$, $\{\gamma\}$ and $\{\lambda_h\}$) and 1 as the initial value for the standard deviation parameters (σ, τ, and $\{\psi_h\}$). In another set of initial values, we changed the setting for the fixed effects by using the estimates from the preliminary analysis of aggregated site-level data. The convergence of the MCMC sampler was assessed by examining trace plots.

The Bayesian estimates of posterior mean, standard deviation, median and 95% credible intervals were summarized for the final model. To visualize the association between each outcome and the latent trait (θ_i), we plot the estimated probability of receiving a process of care (Y_{ijkh}) as a function of the latent trait (θ_i).The parameters characterizing these curves, β_{0h} quantify the intercepts, and λ_h quantifies the steepness of the curve.

$$\Pr(Y_{ijkh}=1)=\frac{\exp(\beta_{0h}+\lambda_h\theta_i)}{1+\exp(\beta_{0h}+\lambda_h\theta_i)} \qquad (3.9)$$

We also estimated the relative contribution of the common variance to total variance of each outcome as $\lambda_h^2/(\lambda_h^2+\psi_h^2)$, where λ_h is the outcome-specific factor loading, and ψ_h^2 is the site-level outcome specific variance. Pairwise Spearman correlation coefficients for the site-level proportion of each individual outcome, the site-level average proportions of the 4 outcomes (AVG), and the estimated latent scores (θ_i) summarize the relationship between observed outcomes and the estimated latent scores. Scatter plots of the latent scores vs. site-level average outcomes are graphed by intervention arm.

3.1.3 Model comparison and fit

We fit other reduced models: one imposing the constraint that the variability in the latent trait (θ_i) across sites is not related to the intervention, but represented by a random component such

that $\theta_i \sim N(0, 1)$; and another model constraints that $\beta_1=0$, because the poor correlation between *oxygen assessment (outcome 1)* and other outcomes suggests this outcome might measure a different underlying construct. The deviance information criterion (DIC) (Spiegelhalter et al., 2002) was computed as an overall measure of model fit to compare models, with smaller DIC being better. The DIC is computed as $\overline{D} + pD$, where $\overline{D}$ is the posterior mean of the deviance and pD is a complexity term that estimates the effective number of parameters in complex hierarchical models, computed as the difference between $\overline{D}$ and the deviance evaluated at the posterior mean of the model parameters.

But does the model fit the data? Global goodness of fit of the final models was assessed using posterior predictive checks (Gelman et al., 1995, Loannis 2009) comparing the observed data with data replicated under the model. Let $y_h^{rep} = (y_{1h}^{rep}, ..., y_{k_h h}^{rep})$ represent the vector of the replicated data for the h^{th} outcome. The distribution of y_h^{rep} given the observed data is

$$p(y_h^{rep} \mid (y_{1h}, ..., y_{k_h h})) = \int p(y_h^{rep} \mid w)\, p(w \mid y_h) d\, w \qquad (3.10)$$

where w is the vector of the model parameters in (3.6). Sampling from (3.10), we replicated 2000 data sets given the model in (3.6). We calculated the empirical distribution for several summary statistics $T_v(y_h)$ for each replicated data set y_h^{rep}, and compared them with the statistics in the observed data set. $T_v(y_h)$ was chosen here as the proportion of patients receiving a recommended process of care by intervention group. The choice of this summary statistic was motivated by the main study interest of evaluating an intervention effect. We report Bayes p-values, estimated using tail area probabilities, by the proportion of times the statistics in the replicated data were more extreme than the observed one, i.e., Bayes p-values= $\Pr(T_v(y_h^{rep}) \geq T_v(y_h) \mid w)$. Bayes p-values were computed for each $T_v(y_h^{rep})$ across intervention

groups. P-values that are close to 0 or 1 are indicative of poor model fit. Values around 0.5 indicate that the distributions of the replicated data and the observed data are close.

3.2 EXTENSION TO THE TWO-LATENT TRAIT MODEL

EDCAP outpatient data and inpatient data with eight outcomes were used to illustrate the proposed method. Because of the structural relationships (structural missing) among outcomes at patient level (e.g., outcomes of outpatient care are measured only on patients who were assigned to outpatient care, and outcomes of inpatient care are collected only for inpatients), let $Y_{ijkh,out}=(y_{ijk1}, y_{ijk2}, y_{ijk3}, y_{ijk4})$ denote a vector of outpatient outcomes, and $Y_{ijkh,in}=(y_{ijk5}, y_{ijk6}, y_{ijk7}, y_{ijk8})$ denote a vector of inpatient outcomes. Please note that each patient has only four outcomes either from outpatient care or inpatient care, due to the structural missing.

Instead of modeling outpatient outcomes and inpatient outcomes by two separate one latent variable models, here we model all the comes in one model framework using two latent trait model, with one trait representing quality of outpatient care and the other one representing quality of inpatient care. Assuming the eight measures all were used in the full model, the covariance structure of the two traits can be assessed.

Within-site Model. Model specification is same as (3.1), but with separate equations for outpatient care and inpatient care due to the structural missing of outcomes at the patient level. Let $\theta_{i,out}, \theta_{i,in}$ denote the outpatient and inpatient traits, respectively. Based on the results from the one latent trait models, different patient level random effects and the same provider effect were specified for the following two equations. The same provider could be admitting some patients and treating others as outpatients.

$$Logit(y_{ijkh,out} \mid \beta_{0h},\beta_{1h},\mu_{ijk,out},\upsilon_{ij},\eta_{ih},z_{ijk}) = \beta_{0h}+\beta_{1h}z_{ijk}+\mu_{ijk,out}+\upsilon_{ij}+\eta_{ih}$$
$$Logit(y_{ijkh,in} \mid \beta_{0h},\beta_{1h},\mu_{ijk,in},\upsilon_{ij},\eta_{ih},z_{ijk}) = \beta_{0h}+\beta_{1h}z_{ijk}+\mu_{ijk,in}+\upsilon_{ij}+\eta_{ih} \quad (3.11)$$

We combine the two equations in (3.11) using an indicator variable of *outp* with 1 indicating that the patient was assigned to outpatient care and 0 indicating that the patient was assigned to inpatient care.

$$Logit(y_{ijkh} \mid \beta_{0h},\beta_{1h},\mu_{ijk,out},\mu_{ijk,in},\upsilon_{ij},\eta_{ih},z_{ijk}) = \beta_{0h}+\beta_{1h}z_{ijk}+(\mu_{ijk,out})^{outp}(\mu_{ijk,in})^{1-outp}+\upsilon_{ij}+\eta_{ih} \quad (3.12)$$

In (3.12), $\mu_{ijk,out} \sim N(0,\sigma_{out}^2)$ and $\mu_{ijk,in} \sim N(0,\sigma_{in}^2)$ denote outpatient and inpatient patient-level random effects respectively, $\upsilon_{ij} \sim N(0,\tau^2)$ is the provider-level random effect, and $\varepsilon_{ih} \sim (0, \psi_h^2)$ is site-level outcome-specific random effect. The site-level random effect (η_{ih}) is specified in between-site model as follows:

Between-Site Model (for extension to two latent traits). Assume there are M traits, here M=2 here. We relate the 8 site-level random effects to a vector (L=2) of latent variables, $\theta_i = (\theta_{1i}, \theta_{2i})$. (3.13)

$$\eta_{ih} \mid \theta_i,\lambda,\Psi_h^2 \sim N(\lambda_{1h}\theta_{1i}+\lambda_{2h}\theta_{2i},\Psi_h^2)$$

Where θ_{1i} *and* θ_{2i} denote the outpatient trait and inpatient trait for site i, respectively.

Here, λ is a 2 X 8 vector of discriminating parameters that account for the correlation among the eight measures. The 8 measurement-specific random effects at the site level were partitioned into two factors, representing the quality of each trait separately, by fixing certain elements of λ equal to 0. The model identification is achieved by anchoring, where we scale each factor by setting one factor loading to specify the ”scale” of the unknown parameters

As in the one latent trait model, variability in a latent quality (θ_{mi}) across sites is assumed to have both a systematic component, explained by site-specific covariates (here the randomized intervention X_i: low, moderate, high intensity intervention), and a random component such that

$$\theta_i = (\theta_{1i}, \theta_{2i}) \sim N_T(X_i`R,\ \Sigma) \tag{3.14}$$

Where $R = \begin{bmatrix} r_{1,out} & r_{2,out} \\ r_{1,in} & r_{2,in} \end{bmatrix}$ is a 2 X 2 vector representing outpatient and inpatient intervention effects of moderate intensity vs. low-intensity, and high-intensity vs. low-intensity interventions, respectively. $\Sigma = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}$ denotes a covariance matrix of the two latent variables.

Formulating the model in equation (3.15) in matrix form, parameters with asterisks are parameters that are fixed at the preassigned values.

$$\begin{bmatrix} \log it(y_{ijk1,out}) \\ \log it(y_{ijk2,out}) \\ \log it(y_{ijk3,out}) \\ \log it(y_{ijk4,out}) \\ \log it(y_{ijk5,in}) \\ \log it(y_{ijk6,in}) \\ \log it(y_{ijk7,in}) \\ \log it(y_{ijk8,in}) \end{bmatrix} = \begin{bmatrix} \beta_{01} \\ \beta_{02} \\ \beta_{03} \\ \beta_{04} \\ \beta_{05} \\ \beta_{06} \\ \beta_{07} \\ \beta_{08} \end{bmatrix} + \begin{bmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{13} \\ \beta_{14} \\ \beta_{15} \\ \beta_{16} \\ \beta_{17} \\ \beta_{18} \end{bmatrix} Z_{ijk} + \begin{pmatrix} \lambda_{11} 0^* \\ 1^* 0^* \\ \lambda_{13} 0^* \\ \lambda_{14} 0^* \\ 0^* \lambda_{24} \\ 0^* 1^* \\ 0^* \lambda_{25} \\ 0^* \lambda_{26} \end{pmatrix} \begin{pmatrix} \theta_{1i} \\ \theta_{2i} \end{pmatrix} + \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \\ \varepsilon_{i5} \\ \varepsilon_{i6} \\ \varepsilon_{i7} \\ \varepsilon_{i8} \end{bmatrix} + (\nu_{ij,in})^{outp_{ijk}} (\nu_{ij,in})^{1-outp_{ijk}} + \mu_{ijk} \tag{3.15}$$

$$\begin{pmatrix} \theta_{1i} \\ \theta_{2i} \end{pmatrix} \sim N\left(\begin{bmatrix} \Upsilon_1 X_i^T \\ \Upsilon_2 X_i^T \end{bmatrix}, \begin{bmatrix} \phi_{11}\phi_{12} \\ \phi_{21}\phi_{22} \end{bmatrix} \right)$$

The augmented complete data likelihood function takes the form:

$$
\begin{aligned}
& f(Y,\Omega \mid \beta_0,\beta_1,\lambda,\gamma,\Sigma,\sigma_{out},\sigma_{in},\tau,\psi,Z,X) \\
&= \prod_{i=1}^{32}\prod_{j=1}^{n_{ij}}\prod_{k=1}^{n_{ijk}}\prod_{h=1}^{h_n} f(y_{ijkh},\theta_i,\mu_{ijk,out},\mu_{ijk,in},\upsilon_{ij},\varepsilon_{ih} \mid \beta_{0h},\beta_{1h},\lambda,\Sigma,\sigma_{out},\sigma_{in},\tau,\psi_h,z_{ijk},X_i) \\
&\propto \prod_{i=1}^{32}\prod_{j=1}^{n_{ij}}\prod_{k=1}^{n_{ijk}}\prod_{h=1}^{h_n}\left(\left(\frac{\exp\{y_{ijkh}(\beta_{0h}+\beta_{1h}z_{ijk}+\mu_{ijk,out}+\upsilon_{ij}+\lambda_h\theta_{i,out}+\varepsilon_{ih})\}}{1+\exp\{y_{ijkh}(\beta_{0h}+\beta_{1h}z_{ijk}+\mu_{ijk,out}+\upsilon_{ij}+\lambda_h\theta_{i,out}+\varepsilon_{ih})\}}\right)^{outp}\times\right. \\
&\quad \left.\left(\frac{\exp\{y_{ijkh}(\beta_{0h}+\beta_{1h}z_{ijk}+\mu_{ijk,in}+\upsilon_{ij}+\lambda_h\theta_{i,in}+\varepsilon_{ih})\}}{1+\exp\{y_{ijkh}(\beta_{0h}+\beta_{1h}z_{ijk}+\mu_{ijk,in}+\upsilon_{ij}+\lambda_h\theta_{i,in}+\varepsilon_{ih})\}}\right)^{1-outp}\right) \\
&\times \prod_{i=1}^{32}\prod_{j=1}^{n_{ij}}\prod_{k=1}^{n_{ijk,out}}\left(\frac{1}{\sigma_{out}}\right)^{\frac{1}{2}}\exp\{\frac{1}{2\sigma_{out}}(\mu_{ijk,out})^2\}\times\prod_{i=1}^{32}\prod_{j=1}^{n_{ij}}\prod_{k=1}^{n_{ijk,in}}\left(\frac{1}{\sigma_{in}}\right)^{\frac{1}{2}}\exp\{\frac{1}{2\sigma_{in}}(\mu_{ijk,in})^2\} \\
&\times \prod_{i=1}^{32}\prod_{j=1}^{n_{ij}}\left(\frac{1}{\tau}\right)^{\frac{1}{2}}\exp\{\frac{1}{2\tau}(\nu_{ij})^2\} \\
&\times \prod_{i=1}^{32}\exp\{-\frac{1}{2}(\theta_i-RX_i)^T\Sigma^{-1}(\theta_i-RX_i)\} \\
&\times \prod_{i=1}^{32}\prod_{h=1}^{n_h}\left(\frac{1}{\psi_h}\right)^{\frac{1}{2}}\exp\{\frac{1}{2\psi_h}(\varepsilon_{ih})^2\}
\end{aligned}
\tag{3.16}
$$

A path diagram is shown in Figure 3-2 for eight outcomes with two latent traits, and four outcomes for each trait.

A Wishart (R, r) prior was specified for the covariance matrix **Ω.** To represent a non-informative prior, we chose a large number of degrees of freedom (r=8). The scale matrix was specified as $R=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$ (Lee and Song 2003). The priors for other parameters were specified as in section 3.1.2.

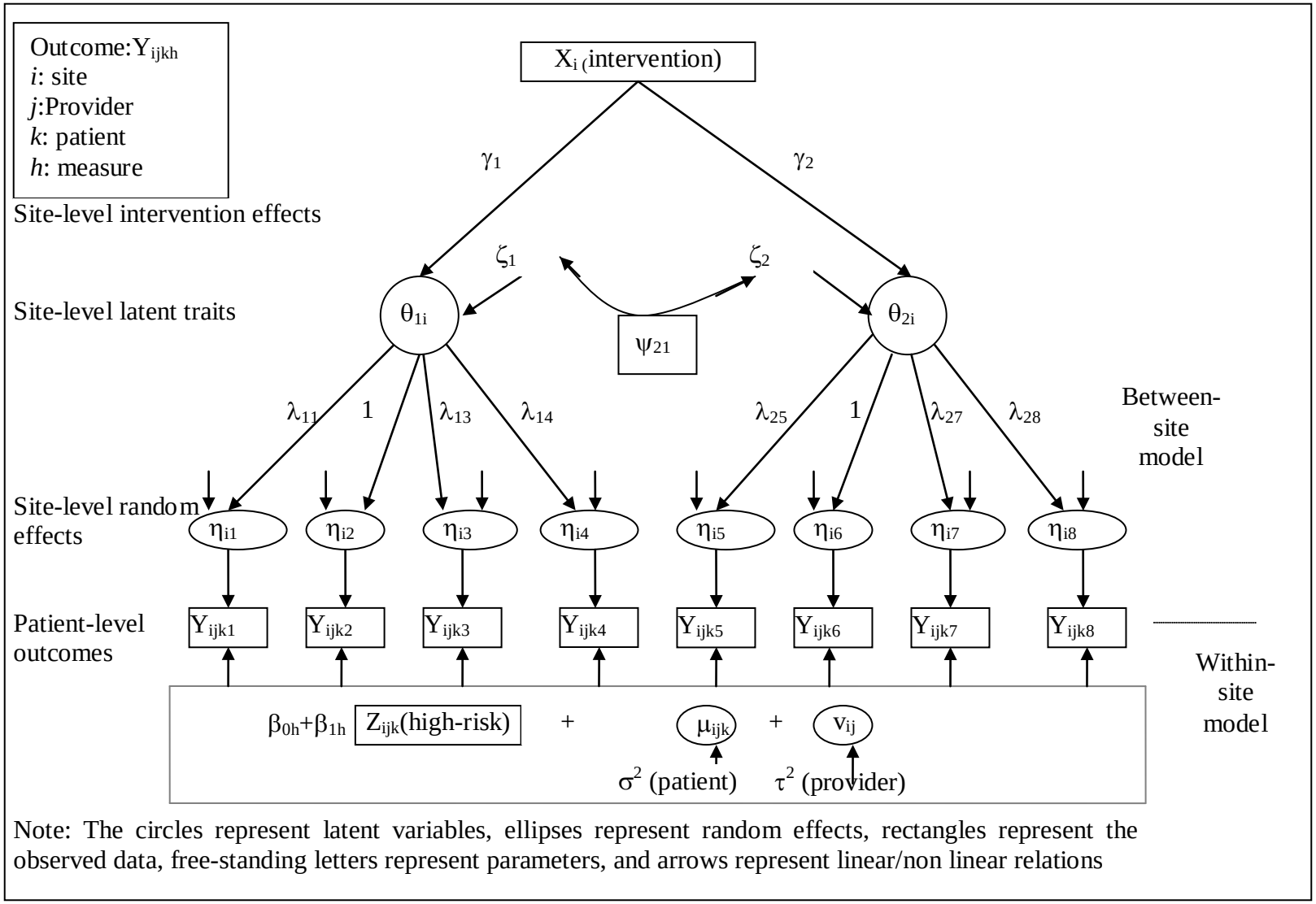


Figure 3-2 Path diagram of two latent trait model with eight outcomes, and four outcomes for each trait

4.0 ONE-LATENT TRAIT MODEL FOR MULTIPLE BINARY OUTCOMES IN A CLUSTER RANDOMIZED CLINICAL TRIAL

4.1 ABSTRACT

In clinical trials, multiple endpoints for treatment efficacy often are obtained, and in addition, data may be collected hierarchically. Statistical analyses become very challenging for such multidimensional hierarchical data, particularly with data collected at more than two levels. We propose a latent variable approach to construct an underlying latent trait from the multiple binary outcomes for a three-level model with multiple outcomes at the patient-level and clustering at the provider and site levels, and assess an intervention effect on the latent trait directly. Random effects model the hierarchical structure, and the parameters are estimated using a fully Bayesian approach. We illustrate the proposed approach in a cluster randomized clinical trial with four binary outcomes. Simulation studies are conducted to check the algorithm and computational implementation. This latent variable modeling approach incorporates the correlation structure into a single latent outcome, and simultaneously regresses the latent outcome on observed covariates. It provides a comprehensive alternative to individual outcomes analysis of multivariate and hierarchical data versus traditional individual outcome analysis, and leads to an intuitively appealing and useful interpretation of complex data.

Key Words: Latent variable models, hierarchical models, Bayesian approach

4.2 INTRODUCTION

In clinical trials, multiple endpoints for treatment efficacy often are obtained (Pocock, Geller, and Tsiatis, 1987). The primary objective of analyses of this kind may not be to identify which particular outcomes differ by groups, but rather to use all the data at hand to establish whether there is a difference between the groups. Several statistical issues arise in evaluating a treatment effect from this type of data. Multidimensionality is a major issue, because several response variables measure slightly different aspects of the effect of interest, and no single response variable suffices as the main outcome variable. Methods that collectively consider all of the response variables are desirable (Gray and Brookmeyer, 1998). When data are collected hierarchically, the outcomes are both multidimensional and hierarchical. These methods also must account for the autocorrelation between observations within the same cluster and the cross-correlation between different response variables both across and within clusters. Frequently, there is a need to summarize all of the multidimensional outcomes into a unidimensional composite score, such as an overall measure of quality of care in health care research (Teixeira-Pinto and Normand 2008).

Commonly used approaches include individual outcome data analysis, dimension reduction, and global test procedures. Each of these approaches fails to characterize relationships between outcomes or summarize those variables. Individual outcome data analysis (i.e., analyze each outcome separately and present multiple P-values) is simple and easy to implement, but fails to provide an overall estimate of the treatment effect, increases the overall Type I error rate, and does not borrow strength across the response variables. Dimension reduction approaches, such as calculating a summary score for several continuous variables, collapsing multiple binary outcomes into a single binary outcome, or constructing a complex function of the individual

response variables such as principal components (Morrison, 1976), may result in biased estimates and a loss of information. Furthermore, data reduction may lead to uninterpretable results, and uncertainty in the aggregated scores is difficult to quantify (Gray and Brookmeyer, 1998). Global testing procedures for multiple outcomes described by O'Brien (1984), Pocock (1987) and Legler (1995) among others, including generalized estimating equation (GEE) based score tests for general M-group comparisons, could be used to assess whether groups differ with respect to multiple outcomes. However such methods do not characterize the relationships between outcomes (Sammel, Ryan and Legler 1997).

Latent variable models (i.e., models that include the random unobserved variables) provide a natural way to analyze such complex multivariate data. (Skrondal and Rabe-Hesketh 2004). The model framework incorporates the correlation structure into a single latent outcome, which is a weighted combination of the observable outcomes, and simultaneously regresses the latent outcome on observed covariates. One advantage of the latent variable model is that a one degree-of-freedom likelihood ratio test of the overall covariate effect could be more powerful than a M degree-of-freedom test (Sammel 1999). Another advantage is that a summary latent score for each individual is estimated as a by-product, such as computing an individual severity score in birth defect study (Legler and Ryan, 1997). These models have been employed extensively in psychological and educational testing (Baker 1992) and the social sciences (Eye and Clogg, 1994). In recent years, use of latent variable models has been increasing in medical and public health research (Legler and Ryan, 1997; Sammel, Ryan, and Legler, 1997; Teixeira-Pinto and Normand, 2008). However, latent variable models seldom have been used to test hypotheses about clinical outcomes in clinical trials and other designed studies (Donaldson 2003).

Most of the literature on latent variable models is frequentist in nature, based on maximum likelihood estimation. Computation is intensive and difficult, because in general no closed form exists for the multi-dimensional integrals. Compared to maximum likelihood methods, Bayesian alternatives using Markov Chain Monte Carlo (MCMC) are becoming more popular and attractive because of their flexibility (Lee and Song 2004, Lee and Song 2003, Dunson et al. 2003). MCMC incorporates prior information and, under specification of a flat (non-informative) prior, provides estimates similar to those obtained by maximum likelihood as well as interval estimates that can be obtained directly from the posterior samples. The difficulties induced by the complexities of the multi-dimensional integrals can be handled efficiently by powerful computing tools, such as the Gibbs sampler. Samples available from the joint posterior distribution of the latent variables can be used flexibly to identify outlying subjects and to obtain important insights into structural relationships (Sik-Yum Lee 2007).

In this paper, we outline a general latent variable model for multiple binary outcomes from a 3-level cluster-randomized clinical trial using a fully Bayesian approach, to evaluate an overall intervention effect with multivariate outcomes at the patient level. In Section 4.3, we present the motivating example of the EDCAP trial (Yealy et al., 2004). In Section 4.4, we describe a Bayesian formulation of the proposed model and the implementation in WinBUGS 1.4 (Windows version of Bayesian inference Using Gibbs Sampling, Gilks 1994) to estimate the parameters. In Section 4.5, we illustrate the proposed methodology using the EDCAP data. We check the algorithm and our computational implementation with simulations in Section 4.6, and conclude with a discussion in section 4.7.

4.3 MOTIVATING EXAMPLE: THE EDCAP STUDY

The cluster-randomized EDCAP study was designed to compare the effectiveness and safety of three guideline implementation strategies of increasing intensity (low-intensity, moderate-intensity, and high-intensity) on quality of care of CAP in EDs (Yealy et al., 2004). The low-intensity implementation arm reflected the quality improvement methods typically used by the collaborating state quality improvement (QI) organizations, which served as a usual care control. The moderate intensity arm also conducted an on-site educational session for ED medical providers and requested a QI plan for the admission decision. The high-intensity intervention also conducted a multifaceted set of provider behavior change techniques (i.e. reminder form, feedback, and bimonthly plan-do-study-act cycles) that continued through the year-long guideline implementation period. The intervention was randomized at the site (ED) level. This study has a 3-level hierarchical data structure, in which multiple outcomes are nested within patients, patients are nested within providers, and providers are nested within clinical sites.

For patients assigned to outpatient care, the study practice guideline recommended four processes of outpatient care in the ED: oxygen assessment, first dose of antibiotics in ED, treatment with compliant antibiotics in the ED, and compliant antibiotic therapy upon discharge. The 1125 patients assigned to outpatient care were seen by 310 providers at 32 clinical sites. Each patient had four binary outcomes, with 1 indicating that the patient received the recommended process of care, 0 otherwise. Scatter plots of the site-level average outcomes (proportions of outpatients receiving recommended processes of care) by intervention arm are shown in Figure 4-1. In general, sites with higher intensity interventions had higher proportions of patients receiving the recommended processes of care. Oxygen assessment had a ceiling

effect, with little variability across sites. Site-level descriptive statistics are summarized in Table 4-1.

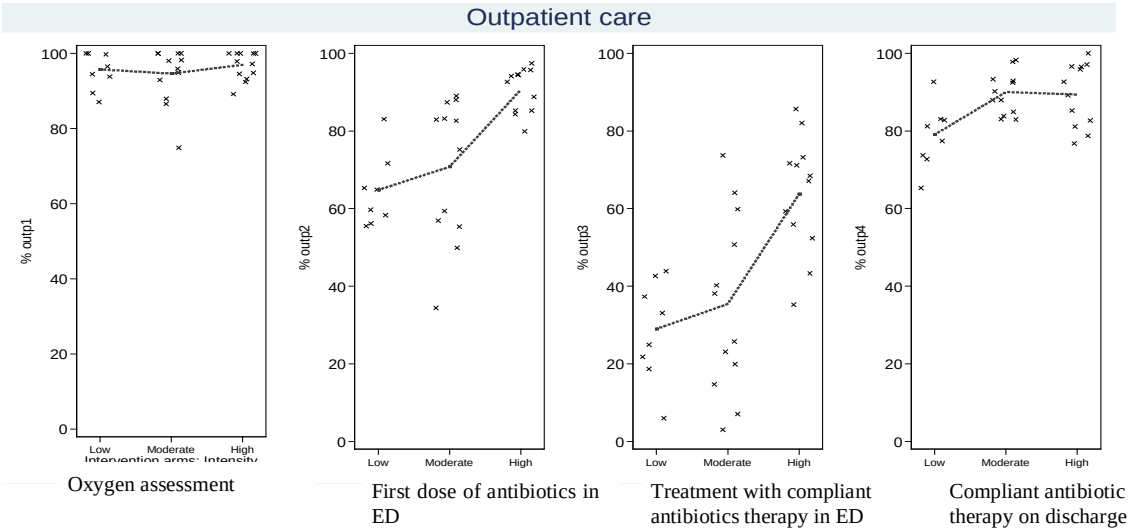


Figure 4-1 Site-level average outcomes (i.e., proportions of outpatients receiving each recommended process of care) by intervention arm

Table 4-1 Site-level averaged proportions of patients receiving each recommended process of care

Variable	Low intensity intervention % N=174	Moderate intensity intervention % N=498	High intensity intervention % N=453
Oxygen assessment	95.7 (4.1)	94.6 (7.5)	96.9 (3.4)
First dose antibiotics in ED	64.8 (9.4)	70.8 (18.2)	90.6 (5.7)
Compliant ED antibiotics	29.0 (13.4)	35.4 (23.1)	63.7 (5.4)
Complaint discharge antibiotics	79.1 (8.9)	90.0 (5.0)	89.4 (8.1)

Estimated intervention effects from an individual outcome analysis are summarized in Table 4-2. We ran separate random effect logistic regression models to estimate the log odds of a patient receiving each recommended process of care as a function of intervention arm, adjusting for patient risk status (low risk vs. high risk) and accounting for clustering at both the provider level and the site level. Models were fit using GLAMM in Stata 10. In total, 12 pairwise

comparisons were conducted for the 4 individual outcomes across 3 intervention arms. Type I error is increased with all of these univariate analyses. Only the intervention effects of the first dose of antibiotics and compliant antibiotic therapy in the ED reached statistical significance once the multiple testing was taken into account using a Bonferroni correction (Miller, 1981). Most importantly, an overall intervention effect cannot be estimated.

Table 4-2 Estimated intervention effects from univariate random effect logistic regression models

Variable	Moderate vs. low intensity intervention Log OR (95% CI)	High vs. low intensity intervention Log OR (95% CI)	High vs. moderate intensity intervention Log OR (95% CI)	Overall P value
Oxygen assessment	0.10 (-1.09, 1.29)	0.40 (-0.82, .63)	0.30 (-0.73, 1.34)	0.78
First dose antibiotic in ED	0.26 (-0.50, 1.01)	**1.90 (1.08, 2.72)**	**1.65 (0.92, 2.37)**	**<0.001**
Compliant ED antibiotic	0.19 (-0.66, 1.04)	**1.78 (0.93, 2.64)**	**1.59 (0.86, 2.33)**	**<0.001**
Complaint discharge antibiotic	**0.77 (0.08, 1.47)**	**0.99 (0.28, 1.70)**	0.22 (-0.41, 0.85)	**0.02**

4.4 LATENT VARIABLE MODEL

We describe our method in the context of EDCAP example. Let Y_{ijkh} (1 = received recommended procedure; 0 = not) be the dichotomous outcome of the h^{th} response (h=1,2,3,4, where the four responses are ordered as: oxygenation, timely antibiotic, compliant ED antibiotic, compliant discharge antibiotic) of the k^{th} (k=1,2,…n_{ij}) patient treated by provider j (j=1,2,…n_i) in the i^{th} (i=1,2…32) site (ED), $X_i\{(x_{i1}, x_{i2})$ *(0,0)*=low intensity, (1,0)=moderate intensity, (1,1)=high intensity)} denote the site-level intervention arms, and Z_{ijk} (0=low risk, 1=high risk) be the covariate of patient-level risk status. The proposed latent approach is a comprehensive, simultaneous model. We formalize the likelihood function by writing the within-site and between-site models sequentially, with the within-site model linking the outcomes to patient risk status and various random effects, and the between-site model linking the site-level random

effect to the latent variable, with the latent mean modeled as a function of the intervention. In this example, the underlying site-level latent variable can be considered as a hypothetical construct of "quality of outpatient pneumonia care in ED".

4.4.1 Model specification

Within-Site Model. Each *o*utcome is assumed to follow a binomial distribution and is linked by logit link to the patient-level covariates, a patient-level random effect, a provider-level random effect, and a site-level outcome-specific random effect:

$$Logit(y_{ijkh} \mid \beta_{0h}, \beta_{1h}, \mu_{ijk}, \upsilon_{ij}, \eta_{ih}, z_{ijk}) = \beta_{0h} + \beta_{1h} z_{ijk} + \mu_{ijk} + \upsilon_{ij} + \eta_{ih} \quad (4.1)$$

In (4.1), β_{0h} denotes a baseline for each outcome, indicating performance of the h^{th} process of care at an average site with the low-intensity intervention for low risk patients, and β_{1h} denote a fixed outcome-specific regression coefficients for a patient-level covariate, risk status (Z_{ijk}). We model the correlation between the four outcomes with a patient-level random effect $\mu_{ijk} \sim N(0, \sigma^2)$, and the correlation between patients within providers with a provider-level random effect $\upsilon_{ij} \sim N(0, \tau^2)$. The η_{ih} denote outcome-specific site level random effects for outcome *h* at site i, permit heterogeneity across the sites, and allow for correlation between patients within a site. The random effects μ_{ijk}, $\nu_{ij,}$ and η_{ih} are assumed to be mutually independent.

Between-Site Model. To model the correlation between the multivariate outcomes at the site-level, we relate the site-level random effects to one latent variable θ_i, such that the four site-level effects are conditionally independent given the latent variable θ_i.

$$\eta_{ih} = \lambda_h \theta_i + \varepsilon_{ih}, \text{ where } \varepsilon_{ih} \sim N(0, \psi_h^2)$$
$$i.e.\ \eta_{ih} \sim N(\lambda_h \theta_i, \psi_h^2) \quad (4.2)$$

Here θ_i serves as a composite profile of outpatient care rendered by site i, with larger values denoting better quality of care. In (4.2), λ_h is fixed outcome-specific discrimination parameter (weight), which quantifies the weight of each outcome on the latent variables and indicates the ability of each outcome to discriminate between sites. Larger values for λ_h correspond to outcomes that better discriminate the underlying quality of care between sites. The sign of λ_h is not identifiable, so the constraint that $\lambda_h > 0$ is added to the model. The $\varepsilon_{ih} \sim (0, \psi_h^2)$ denote site level random error terms for outcome *h,* and quantify the degree of heterogeneity across sites.

In this hierarchical model, variability in latent quality (θ_i) across sites is assumed to have both a systematic component, explained by the site-level interventions (X_i), and a random component such that (4.3)

$$\theta_i = \gamma X_i + \alpha_i \text{ where } \alpha_i \sim N(0,1)$$
$$i.e.\ \theta_i \sim N(\gamma X_i, 1)$$

Where $a_i \sim N(0,1)$ is *i.i.d* error term. The prior variance of the latent quality trait is set to 1, to fix the scale of the latent variable for identifiability of the model (Skrondal & Rabe-Hesketh 2004). Estimates of the latent score ($\hat{\theta}_i$) are given by the posterior mean of θ_i, where low values for the latent score indicate poor quality of care. In (4.3), γ (γ_1 , γ_2) is the vector of fixed treatment effects on latent mean, with γ_1 denoting moderate intensity vs. low intensity, and γ_2 denoting high intensity vs. low intensity respectively.

Equations 4.2 and 4.3 can be combined to formalize the between-site model in one equation:

$$\eta_{ih} = \lambda_h \gamma X_i + \lambda_h \alpha_i + \varepsilon_{ih}, \text{ where } \alpha_i \sim N(0,1), \varepsilon_{ih} \sim N(0, \psi_h^{\,2})$$
$$i.e.\ \eta_{ih} \sim N(\gamma X_i, \lambda_h^{\,2} + \psi_h^{\,2}) \tag{4.4}$$

Combining equations 4.1 and 4.4, the full model can be expanded as:

$$Logit(y_{ijkh} \mid \beta_{0h}, \beta_{1h}, \lambda_h, \theta_i, \mu_{ijk}, \upsilon_{ij}, \varepsilon_{ih}, z_{ijk}) = \beta_{0h} + \beta_{1h} z_{ijk} + \mu_{ijk} + \upsilon_{ij} + \lambda_h \theta_i + \varepsilon_{ih} \tag{4.5}$$

In matrix form

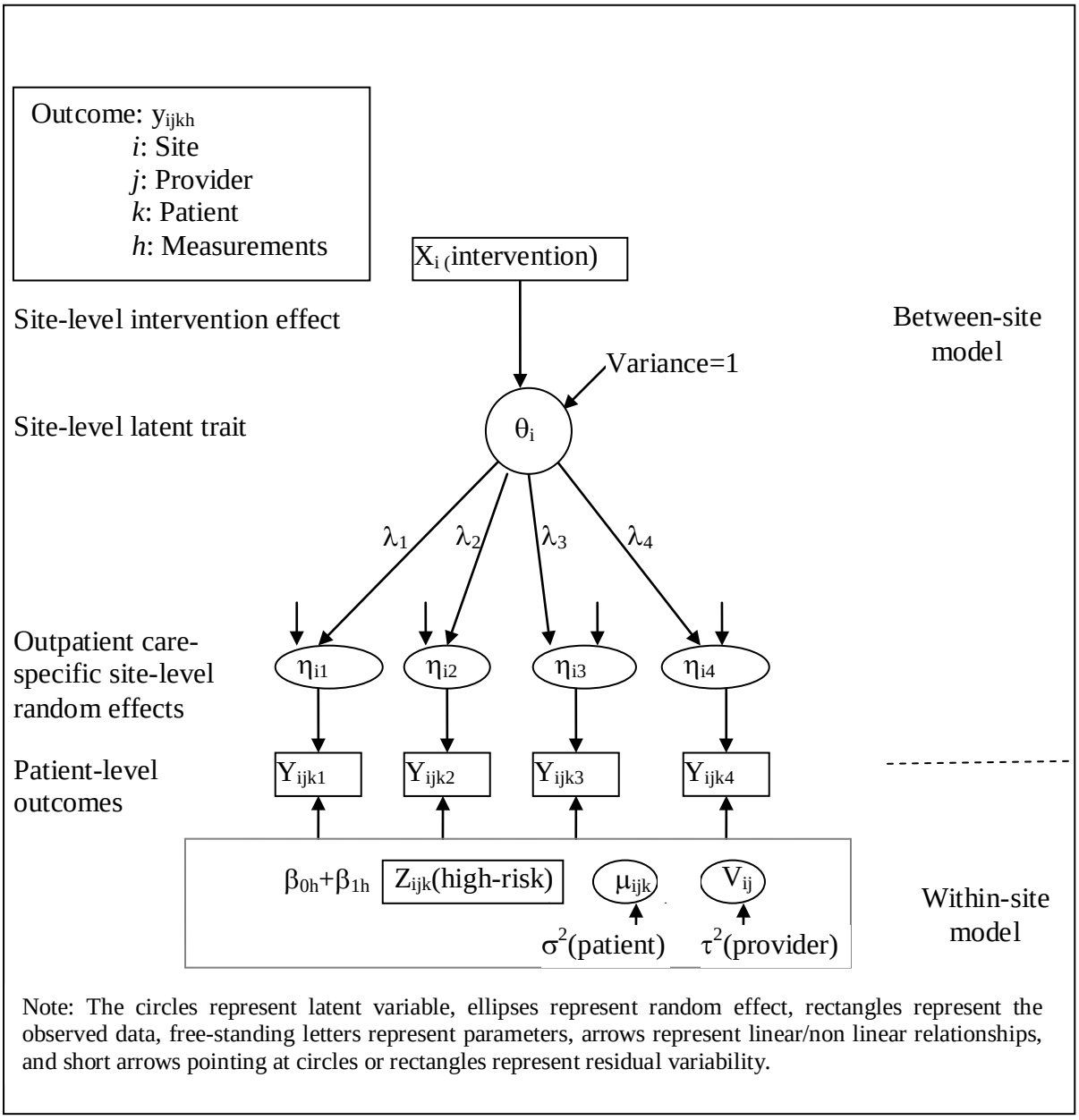
$$\begin{bmatrix} \log it(y_{ijk1}) \\ \log it(y_{ijk2}) \\ \log it(y_{ijk3}) \\ \log it(y_{ijk4}) \end{bmatrix} = \begin{bmatrix} \beta_{01} \\ \beta_{02} \\ \beta_{03} \\ \beta_{04} \end{bmatrix} + \begin{bmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{13} \\ \beta_{14} \end{bmatrix} Z_{ijk} + \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix} [\theta_i] + \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \end{bmatrix} + \nu_{ij} + \mu_{ijk}$$

Wwhere $\theta_i \sim N(\gamma X_i, 1)$ as specified in equation (3.3).

The combined equation can be written as:

$$Logit(y_{ijkh} \mid \beta_{0h}, \beta_{1h}, \lambda_h, \gamma, \mu_{ijk}, \upsilon_{ij}, \varepsilon_{ih}, z_{ijk}, X_i) = \beta_{0h} + \beta_{1h} z_{ijk} + \mu_{ijk} + \upsilon_{ij} + \lambda_h \gamma X_i + \lambda_h \alpha_i + \varepsilon \tag{4.6}$$
$$where\ \mu_{ijk} \sim N(0, \sigma^2), \upsilon_{ij} \sim N(0, \tau^2), \alpha_i \sim N(0,1), \text{ and } \varepsilon_{ih} \sim N(0, \psi_h^{\,2})$$

A total of 20 unknown parameters appear in this model for four outcomes from 3-level hierarchical data, , including the treatment effect $\gamma(\gamma_1, \gamma_2)$, outcome-specific intercept (β_{0h}) , outcome-specific factor loading (λ_h), outcome-specific covariate effect (β_{1h}) , and the variance σ^2 of patient-level random error term (μ_{ijk}), the variance τ^2 of provider-level random error term (v_{ij}) and the outcome- specific variance $\psi_h^{\,2}$ of site-level random error terms (ε_{ih}). To better visualize this complicated model, a path diagram is shown in Figure 4-2.



Note: The circles represent latent variable, ellipses represent random effect, rectangles represent the observed data, free-standing letters represent parameters, arrows represent linear/non linear relationships, and short arrows pointing at circles or rectangles represent residual variability.

Figure 4-2 Path diagram of one latent trait model

4.4.2 Model estimation

Bayesian estimation of the model parameters requires the specification of a prior distribution for each unknown parameter. For the fixed effects ($\{\beta_{0h}\}$, $\{\beta_{1h}\}$ and $\{\gamma\}$), we use diffuse priors, independent normal distributions with mean zero and large variance (10^4), respectively. The prior distribution for factor loading $\{\lambda_h\}$ also was chosen to be N (0, 10^4), but truncated below 0 for identification. For the variance parameters, we followed the recommendation of Gelman (2006) to use a noninformative uniform prior density on standard deviation (SD) parameters in hierarchical models. The uniform (0,100) was used as prior for σ, τ, and $\{\psi_h\}$, the SDs of random effects at the patient, provider and site level, respectively. We assume that the prior distributions of all these parameters are independent, and for each of these parameters, diffuse priors lead to equally probable a priori for all possible values and hence proportional to a constant (Choi K and Seltzer M, 2010). To identify parameters, we fix the scale of the latent variable by setting the variance to 1 for the prior distribution of θ_i.

Bayesian estimates of latent variable models can be obtained using MCMC techniques, including Gibbs sampler (Geman and Geman, 1984) and the Metropolis Hasting (MH) algorithm (Metropolis et al., 1953; Hasting, 1970). The strategy is to treat latent quantities (Ω), including the latent trait and random effects, as hypothetical missing data; due to the nature of MCMC, it is not necessary to integrate out the latent quantities to make inference about the parameters. The latent quantities are updated, along with other parameters from their posterior distributions $P(\theta, \Omega \mid y)$ by drawing observations iteratively from their full conditional densities $p(\theta \mid \Omega, y)$ and $p(\Omega \mid \theta, y)$. For the proposed model, the augmented complete data likelihood function takes the form:

$$f(Y,\Omega \mid \beta_0,\beta_1,\lambda,\gamma,\sigma,\tau,\psi,Z,X)=\prod_{i=1}^{32}\prod_{j=1}^{n_{ij}}\prod_{k=1}^{n_{ijk}}\prod_{h=1}^{4} f(y_{ijkh},\theta_i,\mu_{ijk},\upsilon_{ij},\varepsilon_{ih} \mid \beta_{0\mathrm{h}},\beta_{1\mathrm{h}},\lambda,\sigma,\tau,\psi_h,z_{ijk},X_i)$$

$$=\prod_{i=1}^{32}\left(\prod_{j=1}^{n_{ij}}\left(\prod_{k=1}^{n_{ijk}}\left(\prod_{h=1}^{4} f(y_{ijkh} \mid \beta_{0\mathrm{h}},\beta_{1\mathrm{h}},\lambda,\theta_i,\mu_{ijk},\upsilon_{ij},\varepsilon_{ih})\right)\phi(\mu_{\mathrm{ijk}};0,\sigma)\right)\phi(\upsilon_{\mathrm{ij}};0,\tau)\right)\phi(\theta_i;X_i\,'r,1)\prod_{h=1}^{4}\phi(\varepsilon_{\mathrm{ih}};0,\psi_h)$$

$$\propto\prod_{i=1}^{32}\prod_{j=1}^{n_{ij}}\prod_{k=1}^{n_{ijk}}\prod_{h=1}^{4}\frac{\exp\{y_{ijkh}(\beta_{0\mathrm{h}}+\beta_{1\mathrm{h}}z_{\mathrm{ijk}}+\mu_{\mathrm{ijk}}+\upsilon_{\mathrm{ij}}+\lambda_{\mathrm{h}}\theta_{\mathrm{i}}+\varepsilon_{\mathrm{ih}})\}}{1+\exp\{y_{ijkh}(\beta_{0\mathrm{h}}+\beta_{1\mathrm{h}}z_{\mathrm{ijk}}+\mu_{\mathrm{ijk}}+\upsilon_{\mathrm{ij}}+\lambda_{\mathrm{h}}\theta_{\mathrm{i}}+\varepsilon_{\mathrm{ih}})\}} \qquad (4.7)$$

$$\times\prod_{i=1}^{32}\prod_{j=1}^{n_{ij}}\prod_{k=1}^{n_{ijk}}(\frac{1}{\sigma})^{\frac{1}{2}}\exp\{\frac{1}{2\sigma}(\mu_{\mathrm{ijk}})^2\}\ \times\prod_{i=1}^{32}\prod_{j=1}^{n_{ij}}(\frac{1}{\tau})^{\frac{1}{2}}\exp\{\frac{1}{2\tau}(\nu_{\mathrm{ij}})^2\}$$

$$\times\prod_{i=1}^{32}\exp\{-\frac{1}{2}(\theta_i-\gamma X_i)^2\}\times\prod_{i=1}^{32}\prod_{h=1}^{4}(\frac{1}{\psi_h})^{\frac{1}{2}}\exp\{\frac{1}{2\psi_h}(\varepsilon_{ih})^2\}$$

The generic Bayesian Package WinBUGS1.4 (Spiegelhalter, Thomas, Best and Lunn, 2003) was used to perform MCMC simulations of the posterior distribution. For our model, we used double chains with two sets of varied initial values. The convergence of the MCMC sampler was assessed by examining trace plots.

We fitted two reduced models: one imposing the constraint that the variability in the latent trait (θ_i) across sites is not related to the intervention, but only represented by a random component such that $\theta_i \sim N(0, 1)$; and one imposing the constraints that $\beta_1=0$, because the poor correlation of *oxygen assessment (outcome 1)* with the other outcomes suggests that this outcome might measure a different underlying construct. The deviance information criterion (DIC) (Spiegelhalter et al., 2002) was computed as an overall measure of model fit in model comparison with smaller DIC being better.

The Bayesian estimates of the posterior means, standard deviations, medians and 95% credible intervals were summarized for the final model. To visualize the association between each outcome and the latent trait (θ_i), we plot the estimated probability of receiving a process of

care (Y_{ijkh}) as a function of the latent trait (θ_i). The parameters characterizing these curves, β_{0h} quantify the intercepts, and λ_h quantifies the steepness of the curve.

$$\Pr(Y_{ijkh}=1)=\frac{\exp(\beta_{0h}+\lambda_h\theta_i)}{1+\exp(\beta_{0h}+\lambda_h\theta_i)} \tag{4.9}$$

We also estimated the relative contribution of the common variance to total variance of each outcome as $\lambda_h^2/(\lambda_h^2+\psi_h^2)$, where λ_h is the outcome-specific factor loading, and ψ_h^2 is the site-level outcome specific variance. Pairwise Spearman correlation coefficients for the site-level proportion of each individual outcome, the site-level average proportions of the 4 outcomes, and the latent score (θ_i) summarize the relationship between observed outcomes and the estimated latent score. Scatter plots of the latent score vs. site-level average outcomes are graphed by intervention arm.

Global goodness of fit of the final models was assessed using posterior predictive checks (Gelman et al., 1995, Loannis 2009), comparing the observed data with data replicated under the model. Let $y_h^{rep}=(y_{1h}^{rep},...,y_{k_hh}^{rep})$ represent the vector of the replicated data for the h^{th} outcome. The distribution of y_h^{rep} given the observed data is

$$p(y_h^{rep}\mid(y_{1h},...,y_{k_hh}))=\int p(y_h^{rep}\mid w)\,p(w\mid y_h)d\ w \tag{4.10}$$

Where *w* is the vector of the model parameters in (4.6). Sampling from (4.10), we replicated 2000 data sets given the model in (3.6). We calculated the empirical distribution for several summary statistics $T_v(y_h)$ for each replicated data set y_h^{rep}, and compared them with the statistics in the observed data set. $T_v(y_h)$ was chosen here as the proportion of patients receiving a recommended process of care by intervention group. The choice of this summary statistic was motivated by the main study interest of evaluating intervention effect. We report Bayes p-values,

estimated using tail area probabilities, by the proportion of times the statistics in the replicated data were more extreme than the observed one, i.e. Bayes p-values= $\Pr(T_v(y_h^{rep}) \geq T_v(y_h) \mid w)$. Bayes p-values were computed for each $T_v(y_h^{rep})$ across intervention groups. P-values that are close to 0 or 1 are indicative of poor model fit. Values around 0.5 indicate that the distributions of the replicated data and the observed data are close.

4.5 APPLICATION TO THE EDCAP OUTPATIENT DATA

We applied the proposed approach using outpatient data from the EDCAP trial to assess whether site-level quality of outpatient pneumonia care varied systematically by intervention arm. In total, 20 parameters are estimated in equation 3.6. Our parameter estimates were based on the output of a Gibbs sampler of 7000 iterations after eliminating the first burn-in 3000 iterations (double chains). All the results are based on the pooled two chains, that is, a sample size of 14,000 deviates. The full model with an intervention effect on the latent trait is preferred because the full model clearly shows that the latent trait varies systematically by intervention arm, although the difference in DIC between the two models was small (model with intervention effect: DIC=3211, model without intervention effect: DIC=3215). This is not surprising because the two models differ only in the latent structure at the site level, and only 32 sites are available to test the model fit. The model with "oxygen assessment" removed from the latent trait had slightly higher DIC (3212) than the full model with intervention effect. For comparable interpretation of the four outcomes, we choose to maintain oxygen assessment in the final model.

Posterior summaries of selected parameters characterizing the hierarchical structure and relationships between intervention and outcomes are shown in Table 4-3. The larger the intercept (β_{0h}), the higher the proportion of patients received the therapy on average. *Oxygen assessment* has best average performance, and *treatment with compliant antibiotic therapy in ED* had the poorest average performance. Two outcomes of *first dose of antibiotics in ED* and *treatment with compliant antibiotic therapy in ED* had relatively large values of the discrimination parameter (λ_h), indicating those two therapies had larger variability across sites and potentially more discriminatory power for site performance (and more weight on the latent trait). The data show obvious heterogeneity at patient level (σ=1.37), provider level (τ=0.69) and site level (ψ ranges from 0.43 to 1.38).

The intervention effect was modeled indirectly on the site-level latent trait (θ_i), which could be interpreted as site-level mean quality of ED outpatient pneumonia care. Sites with the high intensity intervention had significantly higher mean quality of outpatient care than sites with the low or moderate intervention (2.78 with 95% posterior intervals (1.21, 4.7) and 2.42 with 95% posterior intervals (1.02, 4.1) respectively (Table 4-3). Latent means between moderate and low intensity intervention arms did not differ significantly. Furthermore, the factor loading parameters λ_h in equation 4.6 also affect the estimation of the direct intervention effect. For comparison with the univariate analyses, global measures of the intervention effects in the scale of log odds ratio are calculated as $X_i\gamma\bar{\lambda}$ (moderate vs. low: 0.18, high vs. low: 1.39 and high vs. moderate: 1.21), which differ somewhat from the average intervention effect on the individual outcomes in Table 4-2 (0.33, 1.27, 0.94, respectively) due to the different weight on each outcome.

Table 4-3 Posterior summaries for the latent variable model

	Parameter	Posterior summaries			
		Mean	Median	sd	95% CI
Baseline (intercept)	β_{01}	4.33	4.32	0.43	(3.53 , 5.23)
	β_{02}	1.09	1.08	0.37	(0.42 , 1.79)
	β_{03}	-1.31	-1.32	0.43	(-2.13 , -0.52)
	β_{04}	2.57	2.58	0.27	(2.03 , 3.08)
Discrimination parameters	λ_1	0.20	0.16	0.16	(0.01 , 0.59)
	λ_2	0.79	0.76	0.19	(0.47 , 1.2)
	λ_3	0.87	0.85	0.24	(0.48 , 1.41)
	λ_4	0.25	0.23	0.14	(0.03 , 0.56)
Stand deviation of random effect					
patient-level	σ	1.37	1.37	0.12	(1.15 , 1.6)
provider-level	τ	0.69	0.69	0.14	(0.4 , 0.95)
Site-level outcome-specific	ψ_1	1.38	1.33	0.44	(0.68 , 2.39)
	ψ_2	0.43	0.43	0.22	(0.06 , 0.89)
	ψ_3	0.68	0.69	0.25	(0.16 , 1.18)
	ψ_4	0.70	0.69	0.21	(0.32 , 1.14)
Intervention effect On Latent mean	γ_1	0.37	0.37	0.58	(-0.77 , 1.48)
	γ_2	2.78	2.72	0.91	(1.21 , 4.7)
	γ_2-γ_1	2.42	2.36	0.79	(1.02 , 4.1)
Global intervention effect on outcomes (Log odds ratio)	$\gamma_1\overline{\lambda}$	0.18	0.19	0.30	(-0.41 , 0.75)
	$\gamma_2\overline{\lambda}$	1.39	1.39	0.34	(0.76 , 2.04)
	$(\gamma_2-\gamma_1)\overline{\lambda}$	1.21	1.20	0.30	(0.64 , 1.82)

Figure 4-3 summarizes the estimated probabilities of receiving a process of care (Y_{ijkh}) as a function of latent trait (θ_i) (equation 3.8). A steeper slope corresponds to higher ability of an outcome to differentiate between sites, and a stronger association with the latent trait. The relationship between the latent trait and outpatient performance outcomes are strongest for *first dose of antibiotics in ED* as well as *treatment with compliant antibiotic therapy in ED*, and

weakest for oxygen assessment. This is not surprising, considering the contribution of the common variance to total variance of each outcome. *First dose of antibiotics in emergency department is* the largest contributor with 77% its total variance explained, followed by *treatment with compliant antibiotic therapy in ED* with 62%, *compliant antibiotic therapy on discharge* with 11% and o*xygen assessment* only 2% of the total variance.

Table 4-4 displays the pairwise Spearman correlation coefficients between the site-level aggregated outcomes and the latent score (θ_i). *Timely first dose in the ED* and *treatment with compliant antibiotic therapy in the ED* are highly correlated (r=0.75) with both the average score and the latent score ($r \geq 0.90$ for each). Oxygen assessment is poorly correlated ($r < 0.3$) with the other variables.

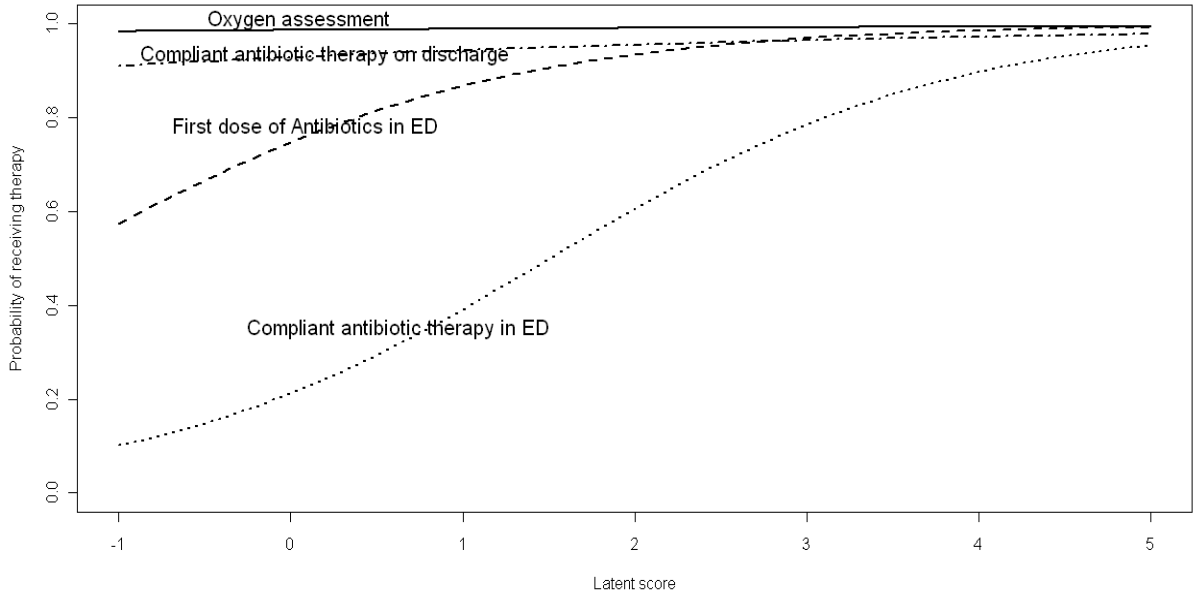


Figure 4-3 Estimated probability of receiving each performance measure as a function of the latent score

Table 4-4 Pairwise Spearman correlation coefficients for site-level aggregated outcomes and the estimated latent score

Variable	**(1)**	(2)	(3)	(4)	AVG	LS
(1) Oxygen assessment	1.00					
(2) First dose antibiotics	0.08	1.00				
(3) Compliant ED antibiotics	-0.14	0.77	1.00			
(4) Complaint discharge antibiotics	-0.15	0.32	0.44	1.00		
(AVG) Site-level average outcomes	0.05	0.90	0.93	0.52	1.00	
(LS) Latent score	0.06	0.92	0.90	0.37	0.96	1.00

Figure 4-4 displays the scatter plot of latent score vs. site-level average outcomes by intervention arm. Higher levels of the intervention generally are associated with higher latent scores, except for a few poorly performing sites in the moderate intensity arm and two relatively well performing sites in the low intensity arm.

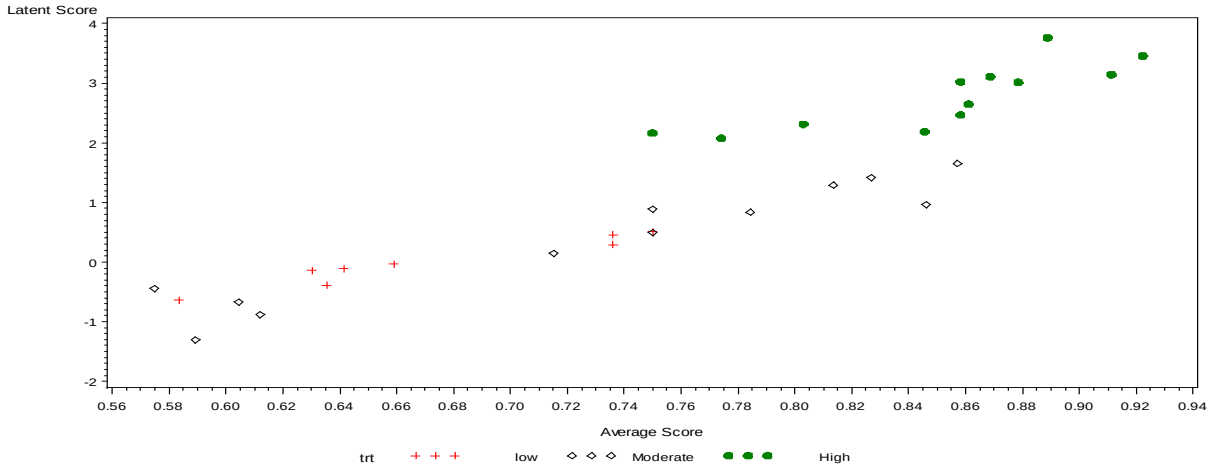


Figure 4-4 Scatter plot of site-level average outcomes and latent score for outpatient data

Table 4-5 summarizes the posterior predictive checks of the fitted models, using the test statistics of site-level average outcomes by intervention arm. The first column T(y) lists the computed proportion from the observed data, and the next column T (y^{rep}) shows the mean and 95% C.I. of the computed proportion from the replicated data. The proposed model estimates

well the percentage of patients receiving each therapy by intervention arm. Most Bayesian p-values are very close to 0.5, and none is close to 1 or 0, indicating no lack of fit of this model.

Table 4-5 Summary of the posterior predictive checks for site-level aggregated outcomes by intervention arm

Variable	Low Intensity Intervention			Moderate intensity Intervention			High Intensity Intervention		
	T(y) %	$T(y^{rep})$ 95% int.	P value	T(y) %	$T(y^{rep})$ 95% int.	P value	T(y) %	$T(y^{rep})$ 95% int.	P value
Oxygen assessment	94.8	94.7 (90.2, 98.3)	0.56	95.6	95.6 (93.2, 97.6)	0.50	96.7	96.7 (94.3, 98.7)	0.51
First dose antibiotics	64.9	65.4 (57.5, 73.6)	0.57	70.1	70.0 (65.5, 74.5)	0.49	91.0	90.8 (87.2, 94.0)	0.47
Compliant ED antibiotics	29.3	28.8 (21.3, 36.2)	0.48	30.7	31 (26.7, 35.5)	0.57	65.6	65.5 (60.3, 70.4)	0.49
Complaint discharge antibiotics	80.5	82.8 (75.9, 88.5)	0.74	89.2	88.3 (84.7, 91.8)	**0**.32	90.7	90.7 (87.4, 93.8)	0.54

4.6 SIMULATION STUDIES

We conducted Monte Carlo simulation studies to check our algorithm and computational implementation of the proposed approach. All computations are performed by iteratively running WinBUGS interface inside SAS following the procedure proposed by Zhang (2008). Although WinBUGS is a convenient tool for estimating Bayesian models, it is not very flexible for simulation studies because it can run only a single model or a single data set at one time. However, SAS can be used to iteratively implement the simulation procedure.

Data were simulated based on the model specified in equation (4.5). The "true" parameter values ranged around parameters estimates obtained from the analysis of outpatients data, to make the population models more realistic. The level-1 unit covariate $X_{ijk,}$ is binary with p=0.5,

and level-4 unit covariate is a three-level categorical variable with p = (1/3,1/3,1/3). The true values were set as: β_0 = (4,1,-1.5,2.5), β_1 = (0.5, 0.2, 1.0, 0.2), σ = 1.0, τ=0.6, ψ= (1.2, 0.5, 0.5, 0.5), λ = (0.2, 0.6, 0.9, 0.3) and γ = (0.5, 2.5). *As* calculated by $\lambda_h^2/(\lambda_h^2+\psi_h^2)$, the fractions of level-4 units variance explained by the common variance are 3%, 59%, 76% and 26% for the four items respectively, which is similar to the outpatient data.

Simulation studies with two different sample sizes were considered, to investigate whether the sample size affected the parameter estimation. In simulation study 1, we generated 30 hospitals with a total of 300 providers and 3000 patients. In simulation study 2, we double the sample size by generating 60 hospitals with a total of 600 providers and 6000 patients. The number of patients per provider varies with an average of 10 patients per provider. For each simulation, we generated 200 datasets. The true parameter values were based on parameter estimates obtained from the EDCAP data, to make the simulation models relevant to this study.

With the prior distribution described in section 4.4.2, we first ran a few sets of data and found that the generated sequences for all parameters converged within 3000 iterations. Although we may use 3000 as the burn-in data, we used a burn-in period of 5000 iterations to ensure the convergence for all the other data sets, and then an additional 5000 observations were collected to produce the Bayesian estimates and their estimated standard errors. Based on the 200 replications, we computed the mean and the SD of the estimates of mean and median, the mean of the SD estimate (mean of SD), as well as the bias (which is the difference between the true parameter and the mean of the corresponding estimates), and the root mean squares (RMS) between the estimates and the true values.

Table 4-6 summarizes the simulation results: (i) As expected, increasing the sample size improves the accuracy of the estimates. (ii) The Bayesian estimates are reasonably accurate, with

bias for almost all parameters being less than 10% in n=3000 and less than 5% in n=6000. The exception was λ_1, whose bias was 57% in the smaller sample and 25% in the larger sample. This is not surprising because only 3% total variance of the first outcome was explained by the latent trait. Also for this reason, the bias of the factor loading λ_1 did not obviously affect the accuracy of the estimates (γ_1, γ_2 and γ_2-γ_1) of the intervention on the latent mean and global intervention ($\gamma_1\bar{\lambda}, \gamma_1\bar{\lambda}$ *and* $(\gamma_1-\gamma_2)\bar{\lambda}$). (iii) In most cases, the estimated SEs and SDs are close to each other; this indicates that the estimated SEs are consistent with the true SEs. (iv) In most cases, the estimates of mean or median are very close, but for the factor loading parameters (λ_1, λ_2, λ_3, λ_4) and the estimates of level-3 units fixed covariate effects (γ_1, γ_2), estimates of the median are closer to the true value with smaller bias in the smaller sample size.

Table 4-6 Posterior summaries for the parameters in the one-latent trait model from simulated data

	N=3000						N=6000					
Parameter	Bias of Mean	SD of Mean	Mean of SE	RMS	Bias of Median	SD of median	Bias of Mean	SD of Mean	Mean of SE	RMS	Bias of Median	SD of median
$\beta_{0,1}$=4	-0.012	0.322	0.374	0.321	-0.017	0.316	0.004	0.219	0.252	0.219	0.002	0.217
$\beta_{0,2}$=1	0.017	0.259	0.265	0.259	0.018	0.26	0.017	0.178	0.18	0.179	0.02	0.178
$\beta_{0,3}$=-1.5	0.005	0.39	0.358	0.39	0.007	0.393	0.015	0.277	0.241	0.276	0.017	0.279
$\beta_{0,4}$=2.5	0.005	0.184	0.214	0.184	0.007	0.183	0.01	0.144	0.144	0.145	0.011	0.144
σ=1.0	0.017	0.071	0.071	0.071	0.017	0.071	0.004	0.048	0.05	0.045	0.004	0.048
ψ_1=1.2	0.103	0.297	0.301	0.315	0.067	0.287	0.051	0.188	0.19	0.195	0.036	0.186
ψ_2=0.5	-0.007	0.124	0.159	0.122	-0.007	0.123	-0.012	0.099	0.108	0.1	-0.009	0.096
ψ_3=0.5	0.014	0.179	0.227	0.179	0.012	0.19	-0.023	0.161	0.173	0.161	-0.017	0.171
ψ_4=0.5	0.004	0.149	0.147	0.148	-0.004	0.146	0.01	0.093	0.095	0.095	0.006	0.091
τ=0.6	0.003	0.062	0.058	0.063	0.002	0.062	0.007	0.042	0.04	0.045	0.007	0.042
λ_1=0.2	0.114	0.169	0.193	0.202	0.083	0.168	0.05	0.111	0.13	0.122	0.037	0.114
λ_2=0.6	0.045	0.163	0.157	0.167	0.032	0.161	0.02	0.104	0.1	0.105	0.014	0.103
λ_3=0.9	0.047	0.194	0.21	0.2	0.033	0.191	0.021	0.137	0.136	0.138	0.017	0.136
λ_4=0.3	0.04	0.125	0.123	0.13	0.029	0.125	0.01	0.083	0.08	0.084	0.006	0.082
γ_1=0.5	0.044	0.612	0.574	0.612	0.027	0.591	-0.004	0.36	0.372	0.359	-0.008	0.357
γ_2=2.5	0.098	1.011	0.868	1.014	-0.01	0.786	-0.015	0.537	0.513	0.536	-0.046	0.517
γ_2-γ_1=2.0	0.055	0.901	0.782	0.901	-0.032	0.689	-0.011	0.485	0.477	0.484	-0.034	0.468
$\gamma_1\overline{\lambda}$ =0.25	0.018	0.296	0.291	0.297	0.017	0.294	0.002	0.184	0.19	0.184	0	0.183
$\gamma_2\overline{\lambda}$ =1.25	0.057	0.264	0.312	0.268	0.048	0.263	0.006	0.218	0.209	0.217	0.001	0.218
$(\gamma_2-\gamma_1)\overline{\lambda}$ =1.0	0.038	0.291	0.312	0.293	0.03	0.29	0.004	0.202	0.206	0.202	0	0.202

4.7 DISCUSSION

Multiple outcomes arise frequently in many fields of biomedical research. Existing commonly used standard software cannot analyze multiple outcomes from hierarchical data with more than two levels. We have described a flexible Bayesian latent variable model for the situation where several binary outcomes are measured to assess an intervention effect in a cluster-randomized clinical trial with 3-level hierarchical data. The difficulties induced by the complexities of the model and the multi-dimensional integrals are handled efficiently by the Gibbs sampler (Geman and Geman, 1984). Compared to traditional separate outcome analyses, this joint modeling approach provides a comprehensive way to analyze multivariate and hierarchical data, and leads to an intuitively appealing and useful interpretation of complex data. In this proposed model framework, the intervention effect was assessed as one degree-of-freedom test, taking advantages of covariance between outcomes. The parameters of factor loading and outcome-specific variance can be used to assess the relationship between outcomes, and identify those outcomes carry the most information about the latent trait. Another advantage of the latent variable model is that it naturally yields a summary measure for each site. As expected, our study shows that the latent scores were highly correlated with raw average outcomes (AVG, r=0.96). As Skrondal and Rabe-Hesketh (2004) pointed out, sometimes an extremely simple approach appears to work as well as much more cumbersome methodologies. However, the raw average methodology cannot be directly applied to clusters with missing data, and cannot incorporate covariance information or relationships between latent variables, in contrast to model based approach.

Methods proposed here easily should be extendable to any distributions from exponential family. For multilevel data, the model could be generalized to allow additional covariates at different hierarchical levels. Covariates can be added to the within-site model (4.1), between-site model (4.2), or on the latent variable level θ_i (4.3) as appropriate. For cross-sectional data without a hierarchical structure, the model can be framed differently by combining the within site and between site models.

In this paper, it has been assumed that the outcomes are conditionally independent given the common latent construct. This bears a close similarity to random-effects models, where, given the random effect, outcomes are assumed to be independent. We will consider extending the model to relax this local independence assumption and allow conditional dependence between outcomes in future work. Another important assumption is that we assume the directions of intervention effects are consistent across outcomes when specifying the prior distribution for factor loadings $\{\lambda_h\}$, as truncated normal distributions above 0 for model identification. This assumption seems reasonable in our EDCAP outpatient data because individual outcome analyses showed the positive intervention effects across all outcomes regardless of significance (Table 4-2). While we considered in this paper the situation where the multiple outcomes are binary, different outcome types, including both continuous and discrete endpoints, can be incorporated into the proposed model framework. Finally, in clinical trials, it might be interesting to examine whether interventions influence the degree of heterogeneity (variation) across health care providers. The estimated variance components would be of great interest, and potentially could be jointly modeled with the regression coefficients.

5.0 TWO-LATENT VARIABLE MODELS FOR ESTIMATING AN INTERVENTION EFFECT FROM MULTIDIMENSIONAL HIERARCHICAL DATA

5.1 ABSTRACT

Multiple outcomes are collected commonly in clinical trials. In this article, we propose fully Bayesian latent variable approaches, including one and two latent variable models, to estimate an overall intervention effect for the situation where multiple outcomes are obtained from hierarchical data. This approach incorporates the correlation structure into one or more latent outcomes, and simultaneously regresses the latent outcome(s) on the intervention. Random effects are included to model the hierarchical structure. This method is applied to data from a cluster-randomized clinical trial with multiple binary outcomes and a 3-level hierarchical structure. Not only does the method allow assessment of intervention effects with respect to multiple outcomes, but it also assesses the relationship between outcomes, identifies those outcomes that carry the most information about the latent trait, and provides a measure of the "quality of care" of each clinical site.

Key Words: hierarchical models, Bayesian approach, multiple latent traits

5.2 INTRODUCTION

Statistical issues that arise when evaluating an intervention effect from multi-center clinical trials with multiple endpoints include multidimensionality (because each response variable measures a slightly different aspect of the effect of interest), autocorrelation between observations within the same site, and cross-correlation between different response variables both across and within sites. Often, a unidimensional composite score is of interest, such as an overall measure of quality of care in health care research (Teixeira-Pinto and Normand 2008).

Commonly used approaches such as dimension reduction and global test procedures fail to characterize relationships between outcomes or summarize their individual effects. Latent variable models (i.e., models that include random unobserved variables) provide a natural way to analyze such complex multivariate data when an underlying hypothetical construct is assumed (Skrondal and Rabe-Hesketh, 2004). Following the very general model framework of generalized linear latent and mixed models (GLLAMMs) (Skrondal & Rabe-Hesketh, 2004 and 2007), latent variable models can be formalized by writing two submodels: a response model and a structural model. The response model specifies the distribution of the observed responses conditional on the latent variables and covariates, which extends generalized linear mixed models (GLMMs) to incorporate factor structure in addition to random intercepts and coefficients; in the structural model, the latent variables themselves may be regressed on other latent and observed covariates.

One advantage of the latent variable model is that a one degree-of-freedom likelihood ratio test of the overall covariate effect can be more powerful than a M degree-of-freedom test (Sammel, 1999). Another advantage is that a summary latent score for each individual can be estimated (Legler and Ryan, 1997). These latent variable models have been employed

extensively in psychological and educational testing (Baker, 1992) and the social sciences (Eye and Clogg, 1994), and in recent years, use has been increasing in medical and public health research (Legler and Ryan, 1997; Sammel, et al., 1997; Donaldson, 2003; Teixeira-Pinto and Normand, 2008). Much of the literature on latent variable models is from a frequentist perspective, based on maximum likelihood estimation. Computation is intensive and difficult, because in general no closed form exists for the multi-dimensional integrals. Bayesian alternatives using Markov Chain Monte Carlo (MCMC) are becoming more popular and attractive because of their flexibility (Lee and Song, 2003 and 2004; Dunson et al., 2003). The difficulties induced by the complexities of the multi-dimensional integrals can be handled efficiently by powerful computing tools, such as the Gibbs sampler (Sik-Yum Lee, 2007).

In this chapter, we implement Bayesian estimation of multilevel latent variable model to asses an overall intervention effect on multiple binary outcomes from a 3-level cluster-randomized clinical trial. In Section 5.3, we present the motivating example of the Emergency Department Community Acquired Pneumonia (EDCAP) study (Yealy et al., 2004), In Section 5.4, we describe our one- and two-latent trait models for clustered binary outcomes, and implement Bayesian estimation in WinBUGS 1.4 (Windows version of Bayesian inference Using Gibbs Sampling, Gilks, 1994). We illustrate the proposed methodology using the EDCAP data in section 5.5, and conclude with a discussion in section 5.6.

5.3 THE EDCAP EXAMPLE

Data from the cluster-randomized EDCAP trial motivated this proposed method. The EDCAP study was designed to compare the effectiveness and safety of three guideline implementation strategies of increasing intensity (low-intensity, moderate-intensity, and high-intensity) on quality of care of Community Acquired Pneumonia (CAP) in Emergency Departments (EDs) (Yealy et al., 2004). The study guideline recommended outpatient care for low risk patients with CAP who presented to the ED, and inpatient care for high risk patients. Four processes of care were recommended for outpatient care and inpatient care respectively, including oxygen assessment, first dose of antibiotics in ED, treatment with compliant antibiotics in the ED, and compliant antibiotic therapy upon discharge for outpatients; oxygen assessment, blood cultures before antibiotic administration, antibiotic administration within 4 hours and treatment with compliant antibiotic therapy in the ED for inpatients. So, there are eight processes of care outcomes considered here.

The EDCAP data have a 3-level hierarchical data structure, in which multiple outcomes are nested within patients, patients are nested within providers, and providers are nested within clinical sites (EDs). The three intervention arms were randomized at the site level. The 3201 patients were seen by 407 providers at 32 clinical sites; 1125 patients were assigned to outpatient care and 2076 were assigned to inpatient care. Each patient had four binary outcomes from either inpatient care or outpatient care, with 1 indicating that the patient received the recommended process of care, 0 otherwise. Site-level average proportions of patients receiving recommended processes of care are summarized in Table 5-1. Scatter plots of the site-level average outcomes (proportions of patients receiving recommended processes of care) by intervention arm are graphed in Figure 5-1. In general, sites with higher intensity interventions

had higher proportions of patients receiving the recommended processes of care, with the exceptions of outpatient and inpatient *oxygen assessment* and *antibiotics within in 4 hours* for inpatients. Oxygen assessment had a ceiling effect, with little variability across sites.

Table 5-1 Site-level average proportions of patients receiving recommended processes of care

Processes of care	Low intensity intervention %	Moderate intensity intervention %	High intensity intervention %
Outpatient processes	N=174	N=498	N=453
Oxygen assessment	95.7 (4.1)	94.6 (7.5)	96.9 (3.4)
First dose antibiotics in ED	64.8 (9.4)	70.8 (18.2)	90.6 (5.7)
Compliant ED antibiotics	29.0 (13.4)	35.4 (23.1)	63.7 (15.4)
Complaint discharge antibiotics	79.1 (8.9)	90.0 (5.0)	89.4 (8.1)
Inpatient processes	N=566	N=661	N=849
Oxygen assessment	97.4 (2.9)	99.2 (1.4)	97.7 (3.2)
Blood cultures before antibiotics	55.8 (17.6)	58.3 (13.0)	74.0 (9.1)
Antibiotics within 4 h	78.9 (14.6)	80.8 (9.2)	77.8 (8.2)
Compliant antibiotics in ED	44.9 (14.2)	59.2 (19.3)	71.2 (14.0)

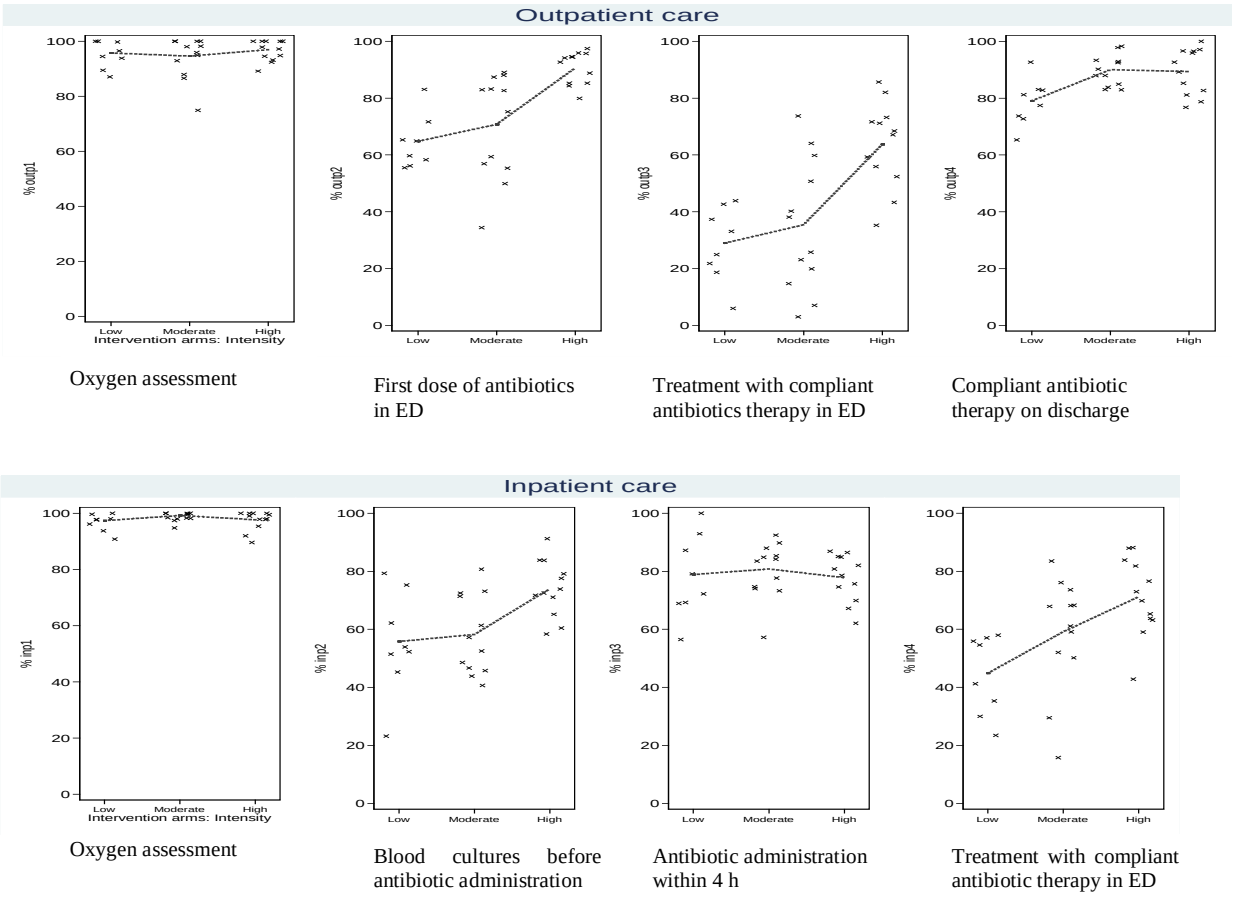


Figure 5-1 Site-level average outcomes (i.e., proportions of patients receiving each recommended process of care) by intervention arm

Estimated intervention effects from analyses of individual outcome are summarized in Table 5-2. We ran separate random effect logistic regression models to estimate the log odds of a patient receiving each recommended process of care as a function of intervention arm, adjusting for patient risk status (low risk vs. high risk) and accounting for clustering at both the provider level and the site level. Models were fitted using GLAMM in Stata 10. In total, 24 pairwise comparisons were conducted for the 8 individual outcomes across 3 intervention arms. Type I error is increased with all of these univariate analyses. Only the intervention effects on outpatient outcomes of the *first dose of antibiotics* and *compliant antibiotic therapy in the ED*, and inpatient outcomes of *blood culture before antibiotics* and *compliant antibiotics in ED* reached statistical significance once the multiple testing was taken into account using a Bonferroni correction

(Miller, 1981). An overall intervention effect cannot be estimated from those individual data analyses.

Table 5-2 Estimated intervention effects from univariate random effect logistic regression models

Variable	Moderate vs. low intensity intervention Log OR (95% CI)	High vs. low intensity intervention Log OR (95% CI)	High vs. moderate intensity intervention Log OR (95% CI)	Overall P value
Outpatient processes of care				
Oxygen assessment	0.10 (-1.09, 1.29)	0.40 (-0.82, .63)	0.30 (-0.73, 1.34)	0.78
First dose antibiotics in ED	0.26 (-0.50, 1.01)	**1.90 (1.08, 2.72)**	**1.65 (0.92, 2.37)**	**<0.001**
Compliant ED antibiotics	0.19 (-0.66, 1.04)	**1.78 (0.93, 2.64)**	**1.59 (0.86, 2.33)**	**<0.001**
Complaint discharge antibiotics	**0.77 (0.08, 1.47)**	**0.99 (0.28, 1.70)**	0.22 (-0.41, 0.85)	**0.02**
Inpatient processes of care				
Oxygen assessment	1.34 (-0.04, 2.73)	2.15 (-0.97, 1.40)	-1.13 (-2.43, 0.17)	0.13
Blood cultures before antibiotics	**0.18 (-0.36, 0.72)**	**0.94 (0.40, 1.48)**	**0.76 (0.27, 1.25)**	**<0.001**
Antibiotics within 4 h	0.11 (-0.47, 0.68)	-0.11 (-0.68, 0.46)	-0.22 (-0.73, 0.30)	0.79
Compliant antibiotics in ED	0.62 (-0.04, 1.29)	**1.27 (0.61, 1.93)**	**0.65 (0.05, 1.24)**	**<0.001**

5.4 LATENT VARIABLE MODELS

We describe our method in the context of EDCAP example. First we analyze the data separately for outpatients and inpatients using one-latent trait models, and then we combine the two sets of data in one model framework using a two-latent trait model to assess the relationship between the two latent traits.

Because of the unobserved structural relationships among outcomes at patient level (e.g., outcomes of outpatient care are measured only on outpatients and outcomes of inpatient care are measured only on inpatients), let $Y_{ijkh,out} = (y_{ijk1}, y_{ijk2}, y_{ijk3}, y_{ijk4})$ denote a vector of outpatient outcomes and $Y_{ijkh,in} = (y_{ijk5}, y_{ijk6}, y_{ijk7}, y_{ijk8})$ denote a vector of inpatient outcomes. Each y_{ijkh} denotes the h^{th} response (ordered as in Table 1) of the k^{th} ($k = 1,2,\ldots n_{ij}$) patient treated by

provider *j* (j = 1,2,…n_i) in site *i* (i = 1,2…32). Let X_i {(x_{i1}, x_{i2}), where *(0, 0)* = low intensity, (1, 0) = moderate intensity, (1, 1) = high intensity intervention)} denote the site-level intervention arms, and Z_{ijk} (0 = low risk, 1 = high risk) be the covariate denoting patient-level risk status. Please note that each patient only has four outcomes from either outpatient care or inpatient care, due to the structural missing data.

5.4.1 One latent trait model

The one latent trait model is illustrated using four outcome measures. Because EDCAP is a cluster randomized clinical trial and the intervention effect is assessed at site level, in the proposed model framework the latent variable is defined at the site level rather than the patient level.

Response Model. The response model regresses the outcomes with a small number of latent variables adjusting for covariates and accounting for the clustering. Each *o*utcome is assumed to follow a binomial distribution and is linked by logit link to a site-level latent trait, a patient-level covariate, a patient-level random effect, a provider-level random effect, and a site-level outcome-specific random effect:

$$Logit(y_{ijkh} \mid \beta_{0h}, \beta_{1h}, \lambda_h, \theta_i \mu_{ijk}, \upsilon_{ij}, \varepsilon_{ih}, z_{ijk}) = \beta_{0h} + \beta_{1h} z_{ijk} + \lambda_h \theta_i + \varepsilon_{ih} + \upsilon_{ij} + \mu_{ijk} \quad (5.1)$$

In matrix form

$$\begin{bmatrix} \log it(y_{ijk1}) \\ \log it(y_{ijk2}) \\ \log it(y_{ijk3}) \\ \log it(y_{ijk4}) \end{bmatrix} = \begin{bmatrix} \beta_{01} \\ \beta_{02} \\ \beta_{03} \\ \beta_{04} \end{bmatrix} + \begin{bmatrix} \beta_{11} \\ \beta_{12} \\ \beta_{13} \\ \beta_{14} \end{bmatrix} Z_{ijk} + \begin{bmatrix} \lambda_1 \\ 1 \\ \lambda_3 \\ \lambda_4 \end{bmatrix} [\theta_i] + \begin{bmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \end{bmatrix} + \nu_{ij} + \mu_{ijk}$$

In (5.1), the β_{0h} denote a baseline for each outcome, indicating performance of the h^{th} process of care at an average site with the low-intensity intervention for low risk patients. The β_{1h} denote fixed outcome-specific regression coefficients for the patient-level covariate of risk status (Z_{ijk}). The term $\lambda_h\theta_i$ models the correlation between the outcomes at the site-level, such that the site-level outcomes are conditionally independent given the latent variable θ_i. In this example, the underlying site-level latent variable θ_i can be considered as a hypothetical construct of "quality of outpatient (or inpatient) pneumonia care in ED" rendered by site i, with larger values denoting better quality of care, and λ_h is fixed outcome-specific discrimination parameter (weight), which quantifies the weight of each outcome on the latent variables and indicates the ability of each outcome to discriminate between sites. Larger values for λ_h correspond to outcomes that better discriminate the underlying quality of care. The model identification is achieved by anchoring, where we set $\lambda_2=1$ to fix the scale of the latent variable. We model the correlation between the four outcomes within patient by a patient-level random effect $\mu_{ijk} \sim N(0,\sigma^2)$, the correlation between patients within providers by a provider-level random effect $\upsilon_{ij} \sim N(0,\tau^2)$, and a site-level outcome-specific random effect $\varepsilon_{ih} \sim (0, \psi_h^2)$ that permits heterogeneity across sites. The random effects μ_{ijk}, $\nu_{ij,}$ and ε_{ih} are assumed to be mutually independent.

Structural Model. The structural model regresses the latent variables on the observed covariate of intervention (X_i), such that variability in latent quality (θ_i) across sites has both a systematic component, explained by the site-level interventions, and a random component:

$$\theta_i = \gamma X_i + \alpha_i \text{ where } \alpha_i \sim N(0,\phi^2), \text{ an i.i.d error term} \quad (5.2)$$
$$\text{i.e. } \theta_i \sim N(\gamma X_i, \phi^2)$$

In (5.2), γ (γ_1 , γ_2) is the vector of fixed treatment effects on the site-level latent trait (θ), with γ_1 denoting a contrast between the moderate intensity vs. low intensity intervention, and γ_2 denoting a contrast between the high intensity vs. low intensity intervention, respectively.

Combining equations 5.1 and 5.2, the full model can be formulated in one equation:

$$Logit(y_{ijkh} \mid \beta_{0h},\beta_{1h},\lambda_h,\gamma,\mu_{ijk},\upsilon_{ij},\varepsilon_{ih},z_{ijk},X_i) = \beta_{0h}+\beta_{1h}z_{ijk}+\lambda_h\gamma X_i+\lambda_h\alpha_i+\mu_{ijk}+\upsilon_{ij}+\varepsilon_{ih} \qquad (5.3)$$

where $\mu_{ijk} \sim N(0,\sigma^2), \upsilon_{ij} \sim N(0,\tau^2), \alpha_i \sim N(0,\phi^2)$, and $\varepsilon_{ih} \sim N(0,\psi_h^{\,2})$

Bayesian estimation of latent variable models treat latent quantities (Ω), including the latent trait and random effects, as hypothetical missing data. For the proposed model, the augmented complete data likelihood function takes the form:

$$f(Y,\Omega \mid \beta_0,\beta_1,\lambda,\gamma,\sigma,\tau,\psi,Z,X) = \prod_{i=1}^{32}\prod_{j=1}^{n_{ij}}\prod_{k=1}^{n_{ijk}}\prod_{h=1}^{4} f(y_{ijkh},\theta_i,\mu_{ijk},\upsilon_{ij},\varepsilon_{ih} \mid \beta_{0h},\beta_{1h},\lambda,\sigma,\tau,\psi_h,z_{ijk},X_i)$$

$$= \prod_{i=1}^{32}\left(\prod_{j=1}^{n_{ij}}\left(\prod_{k=1}^{n_{ijk}}\left(\prod_{h=1}^{4} f(y_{ijkh} \mid \beta_{0h},\beta_{1h},\lambda,\theta_i,\mu_{ijk},\upsilon_{ij},\varepsilon_{ih})\right)\phi(\mu_{ijk};0,\sigma)\right)\phi(\upsilon_{ij};0,\tau)\right)\phi(\theta_i;X_i'r,1)\prod_{h=1}^{4}\phi(\varepsilon_{ih};0,\psi_h)$$

$$\propto \prod_{i=1}^{32}\prod_{j=1}^{n_{ij}}\prod_{k=1}^{n_{ijk}}\prod_{h=1}^{4}\frac{\exp\{y_{ijkh}(\beta_{0h}+\beta_{1h}z_{ijk}+\mu_{ijk}+\upsilon_{ij}+\lambda_h\theta_i+\varepsilon_{ih})\}}{1+\exp\{y_{ijkh}(\beta_{0h}+\beta_{1h}z_{ijk}+\mu_{ijk}+\upsilon_{ij}+\lambda_h\theta_i+\varepsilon_{ih})\}} \qquad (5.4)$$

$$\times\prod_{i=1}^{32}\prod_{j=1}^{n_{ij}}\prod_{k=1}^{n_{ijk}}\left(\frac{1}{\sigma}\right)^{\frac{1}{2}}\exp\left\{\frac{1}{2\sigma}(\mu_{ijk})^2\right\} \times\prod_{i=1}^{32}\prod_{j=1}^{n_{ij}}\left(\frac{1}{\tau}\right)^{\frac{1}{2}}\exp\left\{\frac{1}{2\tau}(\nu_{ij})^2\right\}$$

$$\times\prod_{i=1}^{32}\exp\left\{-\frac{1}{2}(\theta_i-\gamma X_i)^2\right\}\times\prod_{i=1}^{32}\prod_{h=1}^{4}\left(\frac{1}{\psi_h}\right)^{\frac{1}{2}}\exp\left\{\frac{1}{2\psi_h}(\varepsilon_{ih})^2\right\}$$

A total of 20 unknown parameters appear in the full model for four outcomes from 3-level hierarchical data. We also fit reduced models by removing those outcomes that were poorly correlated with other measures or affected by intervention in a different manner, to see if model fit improved. We fitted a reduced model of outpatient data by imposing λ_1=0, since *oxygen assessment (outcome 1)* has a ceiling effect and little variation. We fit a reduced model for inpatient data by imposing λ_1=0 and λ_3=0 for the outcomes of *oxygen assessment* (due to the

ceiling effect) and *antibiotics within 4 hours* because it was not affected by the intervention in the same manner as the other processes of care for inpatients (See Figure 5-1, Table 5-1, Table 5-2).

Path diagrams summarize these full latent trait and reduced models (Figure 5-2). Here circles represent the latent variables, rectangles represent the observed data, free-standing letters represent parameters, arrows connecting circles and/or rectangles represent linear/non linear relationships, and short arrows pointing at circles or rectangles represent residual variability.

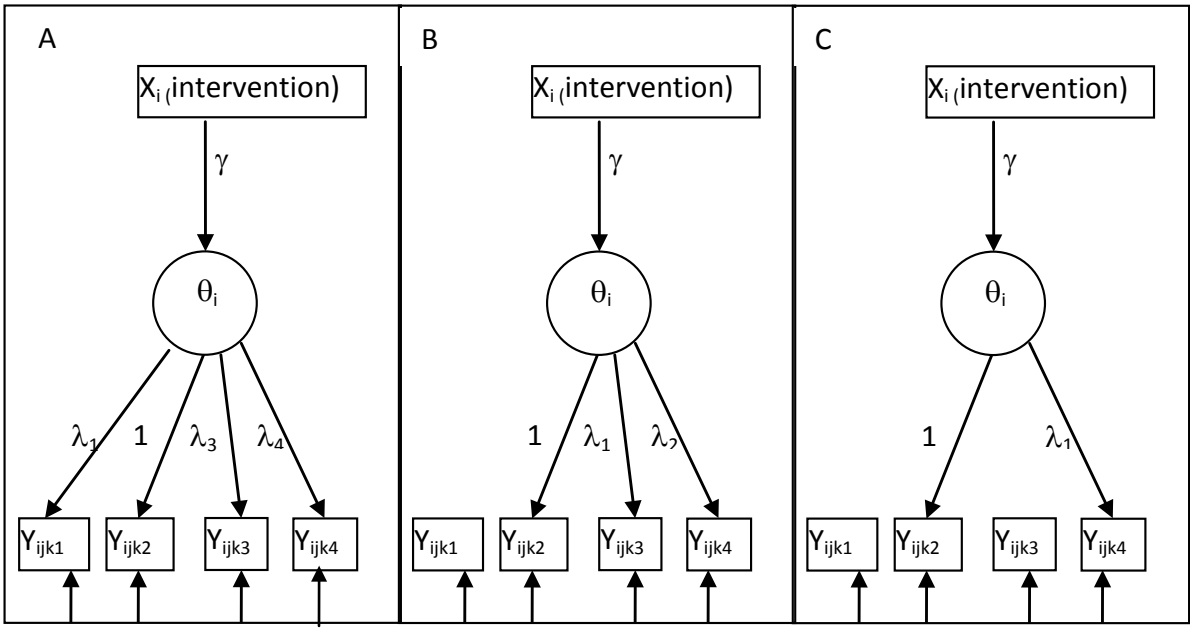


Figure 5-2 Path diagram of one latent trait models. A: full model, B: reduced model for outpatient data, C: reduced model for inpatient data

5.4.2 Two latent trait model

Instead of modeling outpatient outcomes and inpatient outcomes by two separate one latent variable models, here we model all the outcomes in one model framework using a two latent trait model, with one trait representing quality of outpatient care and the other one representing quality of inpatient care. The covariance structure of the two traits also can be assessed.

Response Model. Model specification is same as (3.1), but with separate equations for outpatient care and inpatient care (due to the structural missing of outcomes at the patient level). Let $\theta_{i,out}$ and $\theta_{i,in}$ denote the outpatient and inpatient latent traits, respectively. Based on the preliminary results from the one latent trait models, different patient level random effects and the same provider effect were specified for the following two equations.

$$\begin{aligned} &Logit(y_{ijkh,out} \mid \beta_{0h},\beta_{1h},\lambda_h,\theta_{i,out},\mu_{ijk,out},\upsilon_{ij},\varepsilon_{ih},z_{ijk}) = \beta_{0h}+\beta_{1h}z_{ijk}+\lambda_h\theta_{i,out}+\mu_{ijk,out}+\upsilon_{ij}+\varepsilon_{ih} \\ &Logit(y_{ijkh,in} \mid \beta_{0h},\beta_{1h},\lambda_h,\theta_{i,in},\mu_{ijk,in},\upsilon_{ij},\varepsilon_{ih},z_{ijk}) = \beta_{0h}+\beta_{1h}z_{ijk}+\lambda_h\theta_{i,in}+\mu_{ijk,in}+\upsilon_{ij}+\varepsilon_{ih} \end{aligned} \quad (5.5)$$

We combine the two equations in above by defining an indicator variable of *outp* with 1 indicating that the patient was assigned to outpatient care and 0 indicating that the patient was assigned to inpatient care.

$$\begin{aligned} &Logit(y_{ijkh} \mid \beta_{0h},\beta_{1h},\lambda_h,\theta_{i,out},\theta_{i,in},\mu_{ijk,out},\mu_{ijk,in},\upsilon_{ij},\varepsilon_{ih},z_{ijk}) \\ &\quad = \beta_{0h}+\beta_{1h}z_{ijk}+(\lambda_h\theta_{i,out})^{outp}(\lambda_h\theta_{i,in})^{1-outp}+(\mu_{ijk,out})^{outp}(\mu_{ijk,in})^{1-outp}+\upsilon_{ij}+\varepsilon_{ih} \end{aligned} \quad (5.6)$$

In (3.5), $\mu_{ijk,out} \sim N(0,\sigma_{out}^2)$ and $\mu_{ijk,in} \sim N(0,\sigma_{in}^2)$ denote outpatient and inpatient patient-level random effects respectively, $\upsilon_{ij} \sim N(0,\tau^2)$ denotes provider-level random effects, and $\varepsilon_{ih} \sim$

$(0, \psi_h^2)$ denotes site-level outcome-specific random effects. The model identification is achieved by anchoring, where we fix the scale of each trait by setting one factor loading to 1.

Structural Model. The outpatient and inpatient latent traits are regressed on the intervention separately, and estimates of the intervention effects are allowed to differ across the two traits.

$$\theta_i = (\theta_{i,out}, \theta_{i,in}) \sim N_T(X_i`R,\ \Sigma) \tag{5.7}$$

where $R = \begin{bmatrix} r_{1,out} & r_{2,out} \\ r_{1,in} & r_{2,in} \end{bmatrix}$ is a 2 X 2 vector representing outpatient and inpatient intervention effects of moderate intensity vs. low-intensity and high-intensity vs. low-intensity intervention, respectively, and $\Sigma = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}$ denotes a covariance matrix of the two latent variables. The augmented complete data likelihood function takes the form:

$$\begin{aligned}
&f(Y, \Omega \mid \beta_0, \beta_1, \lambda, \gamma, \Sigma, \sigma_{out}, \sigma_{in}, \tau, \psi, Z, X) \\
&= \prod_{i=1}^{32} \prod_{j=1}^{n_{ij}} \prod_{k=1}^{n_{ijk}} \prod_{h=1}^{h_n} f(y_{ijkh}, \theta_i, \mu_{ijk,out}, \mu_{ijk,in}, \upsilon_{ij}, \varepsilon_{ih} \mid \beta_{0h}, \beta_{1h}, \lambda, \Sigma, \sigma_{out}, \sigma_{in}, \tau, \psi_h, z_{ijk}, X_i) \\
&\propto \prod_{i=1}^{32} \prod_{j=1}^{n_{ij}} \prod_{k=1}^{n_{ijk}} \prod_{h=1}^{h_n} \left(\left(\frac{\exp\{y_{ijkh}(\beta_{0h} + \beta_{1h} z_{ijk} + \mu_{ijk,out} + \upsilon_{ij} + \lambda_h \theta_{i,out} + \varepsilon_{ih})\}}{1 + \exp\{y_{ijkh}(\beta_{0h} + \beta_{1h} z_{ijk} + \mu_{ijk,out} + \upsilon_{ij} + \lambda_h \theta_{i,out} + \varepsilon_{ih})\}} \right)^{outp} \times \right. \\
&\quad \left. \left(\frac{\exp\{y_{ijkh}(\beta_{0h} + \beta_{1h} z_{ijk} + \mu_{ijk,in} + \upsilon_{ij} + \lambda_h \theta_{i,in} + \varepsilon_{ih})\}}{1 + \exp\{y_{ijkh}(\beta_{0h} + \beta_{1h} z_{ijk} + \mu_{ijk,in} + \upsilon_{ij} + \lambda_h \theta_{i,in} + \varepsilon_{ih})\}} \right)^{1-outp} \right) \\
&\times \prod_{i=1}^{32} \prod_{j=1}^{n_{ij}} \prod_{k=1}^{n_{ijk,out}} \left(\frac{1}{\sigma_{out}}\right)^{\frac{1}{2}} \exp\left\{\frac{1}{2\sigma_{out}} (\mu_{ijk,out})^2\right\} \times \prod_{i=1}^{32} \prod_{j=1}^{n_{ij}} \prod_{k=1}^{n_{ijk,in}} \left(\frac{1}{\sigma_{in}}\right)^{\frac{1}{2}} \exp\left\{\frac{1}{2\sigma_{in}} (\mu_{ijk,in})^2\right\} \\
&\times \prod_{i=1}^{32} \prod_{j=1}^{n_{ij}} \left(\frac{1}{\tau}\right)^{\frac{1}{2}} \exp\left\{\frac{1}{2\tau} (\nu_{ij})^2\right\} \\
&\times \prod_{i=1}^{32} \exp\left\{-\frac{1}{2} (\theta_i - RX_i)^T \Sigma^{-1} (\theta_i - RX_i)\right\} \\
&\times \prod_{i=1}^{32} \prod_{h=1}^{n_h} \left(\frac{1}{\psi_h}\right)^{\frac{1}{2}} \exp\left\{\frac{1}{2\psi_h} (\varepsilon_{ih})^2\right\}
\end{aligned} \tag{5.8}$$

Preliminary results of the one latent trait models showed that the outpatient latent trait was constructed mainly by three outcomes (*first dose of antibiotics in ED, compliant ED antibiotics and complaint discharge antibiotics*), and the inpatient latent trait was constructed mainly by two outcomes (*Blood cultures before antibiotics* and *Compliant antibiotics in ED)*. These five outcomes are used to fit the two-latent trait model. Though two separate latent traits were modeled, this integrated model will allow exploration of the relationship between the two latent traits. In the path diagram shown in Figure 5-3, ϕ_{12} denotes the correlation between the latent traits for outpatient and inpatient care conditional on intervention effects.

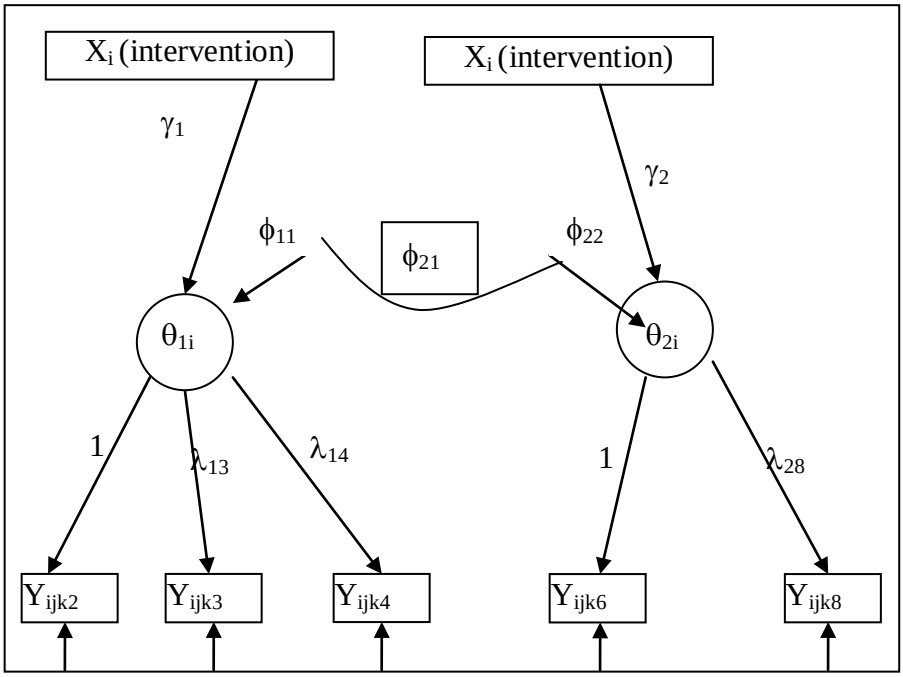


Figure 5-3 Path diagram of the two latent traits model

5.4.3 Model estimation

Bayesian estimation of the model parameters requires the specification of a prior distribution for each unknown parameter. For the fixed effects ($\{\beta_{0h}\}$, $\{\beta_{1h}\}$, $\{\gamma\}$) and factor loadings $\{\lambda_h\}$, we use diffuse priors N (0, 10^4), independent normal distributions with mean zero and large variance, respectively. For the variance parameters (σ, τ, $\{\psi_h\}$ and ϕ), we used uniform (0,100) as priors (Gelman 2006). In the two latent traits model, a Wishart (R, r) prior was specified for the covariance matrix Ω**.** To represent non-informative prior, we chose a large degree of freedom as r=8. The scale matrix was specified as $R = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ (Lee and Song 2003). We assume that the prior distributions of each these parameters are independent. The generic Bayesian Package WinBUGS1.4 (Spiegelhalter, Thomas, Best and Lunn, 2003) was used to simulate the posterior distribution using MCMC. For our model, we used double chains with two sets of varied initial values. The convergence of the MCMC sampler was assessed by examining trace plots. The deviance information criterion (DIC) (Spiegelhalter et al., 2002) was computed as an overall measure of relative model fit to compare full and reduced models, with smaller DIC being better.

The Bayesian estimates of posterior means and 95% credible intervals are summarized for the final models. Estimates of the site specific latent scores (i.e., value of latent variable at each site) are given by the posterior mean of θ_i. To visualize the association between each outcome and the latent trait ((θ_i), we plot the estimated probability of receiving a process of care (Y_{ijkh}) as a function of the latent trait (θ_i). Of the parameters charactering these curves, β_{0h} quantifies the intercept, and λ_h quantifies the steepness of the curve.

$$\Pr(Y_{ijkh}=1)=\frac{\exp(\beta_{0h}+\lambda_h\theta_i)}{1+\exp(\beta_{0h}+\lambda_h\theta_i)} \tag{5.9}$$

We also estimated the relative contribution of the common variance to the total variance of each outcome as

$$\lambda_h^2\phi^2/(\lambda_h^2\phi^2+\psi_h^2) \tag{5.10}$$

IN (5.10), λ_h is the outcome-specific factor loading, and ψ_h^2 is the site-level outcome specific variance. Scatter plots of the outpatient latent score vs. inpatient latent score from the two-latent trait model are shown by intervention arm.

5.5 APPLICATION TO THE EDCAP DATA

We applied the proposed approaches to the EDCAP data, to assess whether site-level quality of pneumonia care varied systematically by intervention arm. One latent trait models were fit first for outpatient data and inpatient data separately, and then the two latent trait model was fit to model inpatient and outpatient care simultaneously. Our parameter estimates were based on the output of a Gibbs sampler of 10,000 iterations after eliminating the first burn-in of 10000 iterations (double chains). All the results are based on the pooled two chains, that is, a sample size of 20,000 deviates.

5.5.1 One latent trait model results

Compared to the respective full models (Figure 5-2a), both the outpatient reduced model with *oxygen assessment* removed from the latent trait (Figure 5-2b) and inpatient reduced model with

oxygen assessment and antibiotics within 4 hours (Figure 5-2c) had almost the same DIC (Outpatient full model vs. reduced model: 3207, 3208; Inpatient full model vs. reduced model: 7524, 7523). Given the tiny differences in DIC values between the full model and reduced model, we choose to present here the results from the full model. Results of the reduced models are not shown, but the corresponding estimates of the parameters are very similar to the full models shown here.

Posterior summaries of selected parameters characterizing the hierarchical structure and relationships between intervention and outcomes of the full models are shown in Table 5-3. The larger the intercept (β_{0h}), the higher the proportion of patients received the therapy on average. *Oxygen assessment* has best average performance, which is in agreement with ceiling effect seen in Figure 1, and *treatment with compliant antibiotic therapy in outpatient care* had the poorest average performance. Two outcomes of outpatient care (*first dose of antibiotics in ED* and *treatment with compliant antibiotic therapy in ED*) and two outcomes of inpatient care (*blood culture before antibiotics and treatment with compliant antibiotic therapy in ED*) had relatively larger values of the discrimination parameter (λ_h); larger weight on the latent traits indicates that those therapies had larger variability across sites and potentially more discriminatory power for site performance. The data show obvious heterogeneity at patient-level (σ = 1.36 and 0.50 for outpatients and inpatients, respectively), provider-level (τ=0.70 and 0.43) and site-level (ψ ranges from 0.48 to 1.43).

The intervention effect was modeled on the site-level latent trait (θ_i), which could be interpreted as a site-level mean quality of ED outpatient (or inpatient) pneumonia care. Sites with the high intensity intervention had significantly higher mean quality of outpatient and inpatient care than sites with the low or moderate intensity interventions (outpatient: estimate of mean

differences of high vs. low = 1.98, 95% posterior interval = (0.98, 3.06) and high vs. moderate = 1.73, 95% posterior interval = (0.85, 2.63) respectively; inpatient: high vs. low = 0.72, 95% posterior interval = (0.14, 1.33) and high vs. moderate = 0.50, 95% posterior interval = (0.06, 1.1) respectively, Table 5-3). Latent means between moderate and low intensity intervention arms did not differ significantly.

To estimate the direct overall intervention effect, we must recall that the factor loading parameters λ_h in equation 3.3 also affect the estimation of the direct intervention effect. Hence, global measures of the intervention effects in the scale of log odds ratio are calculated as $X_i \gamma \bar{\lambda}$ (Outpatient: 0.16, 1.21 and 1.06 for moderate vs. low, high vs. low and high vs. moderate respectively; inpatient: 0.14, 0.36 and 0.23 respectively). None of the global intervention effects is significant for inpatient data, because the two outcomes of *oxygen assessment* and *antibiotics within 4 hours* are negatively associated with the intervention by negative factor loadings (-0.48, -0.28 respectively), which is in agreement with the individual outcome analyses in Table 5-2.

Table 5-3 Posterior summaries for one latent variable models of outpatient and inpatient data

	Parameter	Outpatient model Mean	95% CI	Inpatient model Mean	95% CI
Baseline (intercept)	β_{01}	4.58	(3.64,5.73)	4.42	(3.3 , 5.72)
	β_{02}	1.13	(0.34,1.89)	0.16	(-0.25 , 0.59)
	β_{03}	-1.30	(-2.26,-0.47)	1.41	(1 , 1.87)
	β_{04}	2.59	(2,3.11)	-0.10	(-0.65 , 0.45)
Discrimination parameters	λ_1	0.00	(-0.7,0.64)	-0.48	(-4.75 , 2.79)
	λ_2	1.00*	NA	1.00*	NA
	λ_3	1.18	(0.76,1.75)	-0.28	(-2.04 , 0.84)
	λ_4	0.31	(-0.04,0.69)	1.99	(0.43 , 6.41)
Stand deviation of random effect					
patient-level	σ	1.36	(1.13,1.6)	0.50	(0.32 , 0.66)
provider-level	τ	0.70	(0.46,0.97)	0.43	(0.32 , 0.53)
Site level outcome-specific	ψ_1	1.40	(0.64,2.45)	1.43	(0.7 , 2.54)
	ψ_2	0.48	(0.08,0.92)	0.58	(0.35 , 0.85)
	ψ_3	0.63	(0.1,1.16)	0.60	(0.39 , 0.87)
	ψ_4	0.70	(0.29,1.16)	0.72	(0.32 , 1.05)
Stand deviation of latent trait	ϕ	0.75	(0.39,1.16)	0.14	(0.01 , 0.42)
Intervention effect on latent mean	γ_1	0.25	(-0.61,1.19)	0.22	(-0.2 , 0.65)
	γ_2	**1.98**	**(0.98,3.06)**	**0.72**	**(0.14 , 1.33)**
	γ_2-γ_1	**1.73**	**(0.85,2.63)**	**0.50**	**(0.06 , 1.1)**
Global intervention effect (Log odds ratio)	$\gamma_1\overline{\lambda}$	0.16	(-0.39,0.76)	0.14	(-0.06 , 0.53)
	$\gamma_2\overline{\lambda}$	**1.21**	**(0.56,1.99)**	0.36	(-0.13 , 0.95)
	$(\gamma_2-\gamma_1)\overline{\lambda}$	**1.06**	**(0.49,1.7)**	0.23	(-0.13 , 0.61)

*fixed values for model identification

Figure 5-4 displays the estimated probability of receiving a recommended process of care (Y_{ijkh}) as a function of latent trait (θ_i) (equation 5.7). A steeper slope corresponds to higher ability of an outcome to differentiate between sites, and a stronger association with the latent trait. The relationship between the latent trait and outpatient performance outcomes are strongest for *first dose of antibiotics in ED* as well as *treatment with compliant antibiotic therapy in ED,*

and weakest for oxygen assessment. The relationship between the latent trait and inpatient performance outcomes are strongest for *blood culture before antibiotics* as well as *treatment with compliant antibiotic therapy on discharge*, and weakest for the other two outcomes, even negatively associated with *antibiotics within in 4 hours*. This result is not surprising, as looking at the contribution of the common variance to total variance of each outcome calculated by equation (5.10). The values are 0%, 71%, 66% and 10% for the four outpatient outcomes, respectively, and 0.2%, 6%, 0.4% and 13% for the four inpatient outcomes respectively (outcomes are ordered as in Table 5-1,.

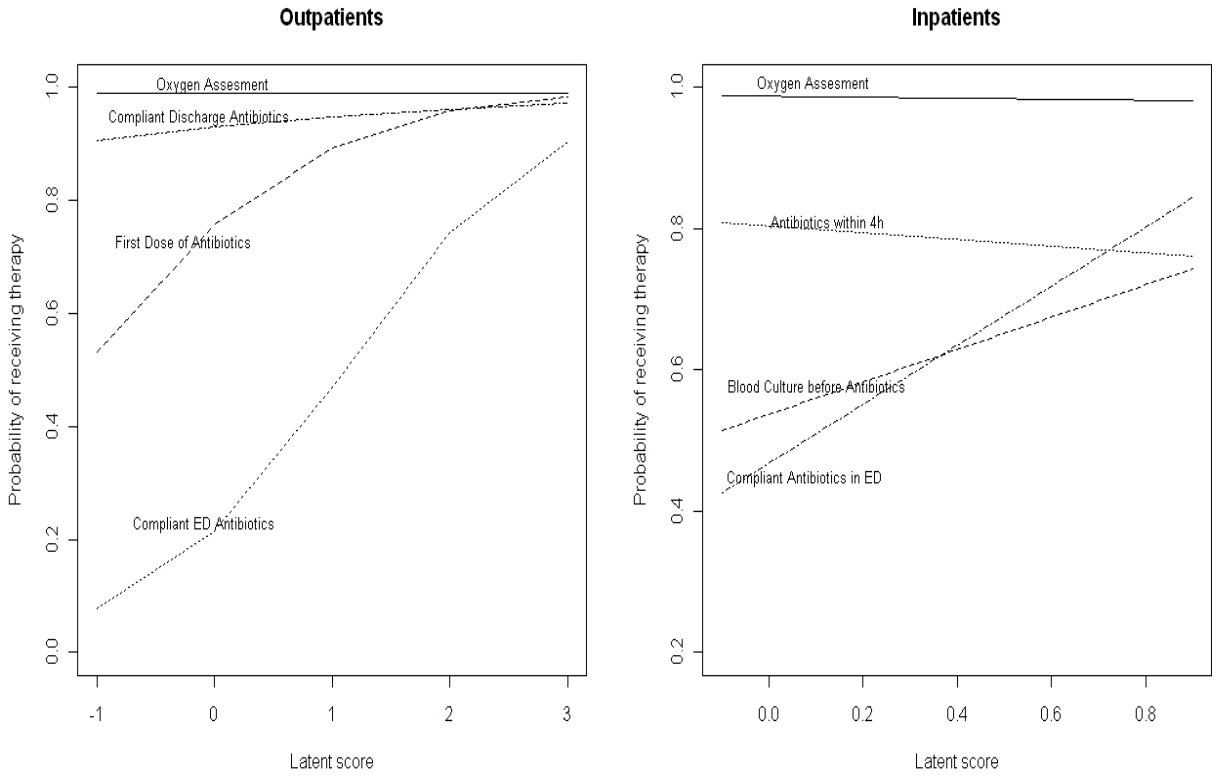


Figure 5-4 Estimated probability of receiving each performance measure as a function of the latent score for outpatients and inpatients

5.5.2 Two latent trait model results

Posterior summaries of selected parameters in the two latent trait model characterizing the latent traits and intervention effects are shown in Table 5-4. The point estimates of discrimination parameters and intervention effects for the outpatient outcomes are very close to the results from the one-latent trait model, but with a stronger intervention effect. The global intervention effect in inpatient care turned to be significant because the two outcomes, oxygen assessment that was not associated and antibiotics within 4 hours that was negatively associated with the interventions were removed from the analysis, otherwise, the global intervention effect will be cancelled out. This is analogous to interaction data analysis, i.e., if the intervention effect significantly varies across outcomes, it is not appropriate to estimate overall single intervention effect on all outcomes. As in the single latent trait models, no intervention effect was observed for the moderate intensity versus low intensity comparison. The intervention effects were stronger in outpatient care than inpatient care. The outpatient latent trait had higher variation than inpatient latent trait (variance of 0.24 vs. 0.11), and the two traits are uncorrelated after accounting for the intervention effect (r = 0.06, 95% posterior interval: -0.53, 0.61).

Figure 5-5 displays the scatter plot of estimated outpatient latent score vs. inpatient latent score by intervention arm. For both traits, higher levels of the intervention generally are associated with higher latent scores, except for a few poorly performing sites in the moderate intensity arm. Overall, the two traits are positively correlated, but within in each intervention arm, there seems to be no correlation, i.e., the two latent traits do not exhibit site-level clustering after adjusting for intervention arm.

Table 5-4 Posterior summaries of the two latent trait model

	Parameter	Posterior summaries			
		Mean	Median	sd	95% CI
Discrimination parameters	λ_{13}	1.13	1.11	0.27	(0.67 , 1.71)
	λ_{14}	0.33	0.32	0.20	(-0.03 , 0.74)
	λ_{28}	0.92	0.90	0.37	(0.26 , 1.71)
Intervention effect on latent mean, outpatient	γ_1	0.40	0.38	0.45	(-0.44 , 1.31)
	γ_2	**2.37**	**2.35**	**0.53**	**(1.36 , 3.47)**
	γ_2-γ_1	**1.97**	**1.96**	**0.43**	**(1.15 , 2.83)**
Global intervention effect on outcomes, outpatient	$\gamma_1\overline{\lambda}$	0.32	0.31	0.36	(-0.36 , 1.08)
	$\gamma_2\overline{\lambda}$	**1.92**	**1.91**	**0.40**	**(1.18 , 2.73)**
	$(\gamma_2-\gamma_1)\overline{\lambda}$	**1.59**	**1.59**	**0.32**	**(0.98 , 2.23)**
Intervention effect on latent mean, inpatient	γ_1	0.25	0.24	0.25	(-0.22 , 0.76)
	γ_2	**0.98**	**0.97**	**0.27**	**(0.49 , 1.55)**
	γ_2-γ_1	**0.74**	**0.73**	**0.27**	**(0.22 , 1.29)**
Global intervention effect on outcomes, inpatient	$\gamma_1\overline{\lambda}$	0.24	0.23	0.23	(-0.19 , 0.72)
	$\gamma_2\overline{\lambda}$	**0.93**	**0.93**	**0.25**	**(0.44 , 1.45)**
	$(\gamma_2-\gamma_1)\overline{\lambda}$	**0.69**	**0.69**	**0.25**	**(0.21 , 1.19)**
Covariance between latent traits	Φ_{11}	0.24	0.21	0.14	(0.07 , 0.6)
	Φ_{22}	0.11	0.10	0.05	(0.05 , 0.23)
	Φ_{12}	0.01	0.01	0.06	(-0.1 , 0.14)
	corr	0.06	0.07	0.30	(-0.53 , 0.61)

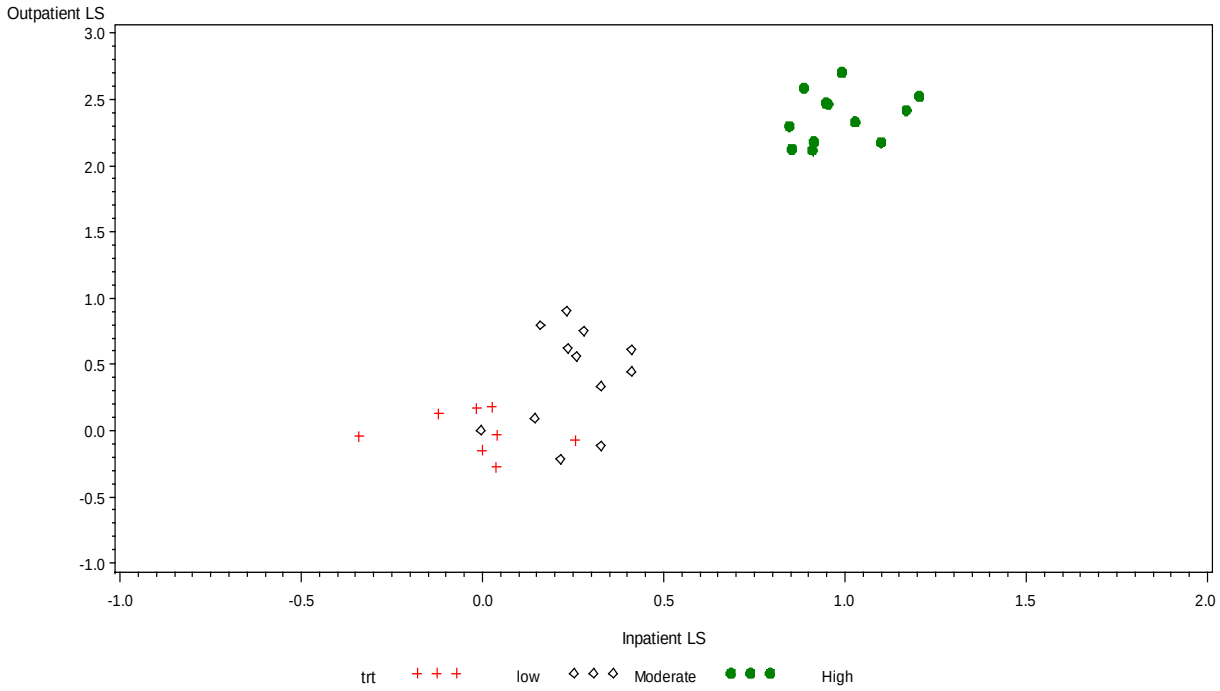


Figure 5-5 Scatter plot of outpatient latent scores versus inpatient latent scores

5.6 DISCUSSION

We have described a flexible Bayesian latent variable model for the situation where several binary outcomes are measured to assess an intervention effect in a cluster-randomized clinical trial with 3-level hierarchical data. The method not only allows assessment of intervention effects with respect to multiple outcomes, but also assesses the relationship between outcomes, identifies those outcomes that carry the most information about the latent trait(s), allows for structural missingness, and provides a summary measure of the "quality of care" at each clinical site across patient subgroups. In the EDCAP study, this approach showed that latent scores for the quality of both outpatient and inpatient care were highest for the high intensity intervention sites, and that these traits were essentially uncorrelated.

Methods proposed here are extendable to any distributions of exponential family. For multilevel data, the model can be generalized to allow additional covariates at either response model (first stage) or structural model (second stage), as appropriate. In EDCAP study, the intervention was added at the second stage because the intervention was randomized at the site level.

One important assumption is that the outcomes are conditionally independent given the common latent construct. This bears a close similarity to random-effects models, where outcomes are assumed to be independent given the random effects. We will consider extending the model to relax this local independence assumption to allow conditional dependence between outcomes in future work. While we considered multiple binary outcomes in this paper, different outcome types can be incorporated into the proposed model framework, including both continuous and discrete endpoints.

In summary, this latent variable approach provides a comprehensive alternative to traditional individual outcome analysis to quantify intervention (or exposure) effects with regard to multiple outcomes in hierarchical data setting. The two latent trait model allows a joint assessment of quality of care for two distinct subgroups of patents.

6.0 CONCLUSION AND DISCUSSION

In clinical trials, multiple endpoints for treatment efficacy often are obtained (Pocock, Geller, and Tsiatis, 1987), and in addition, data may be collected hierarchically. Commonly used approaches, such as individual outcome data analysis, dimension reduction, or global test procedures, fail to borrow strength across outcomes, characterize relationships between outcomes or summarize those variables. Latent variable approaches provide a natural way to analyze complex multivariate hierarchical data, but seldom have been used to test hypotheses about clinical outcomes in clinical trial and other designed studies (Donaldson 2003). In addition, no existing commonly used software could analyze multivariate outcomes from hierarchical data with more than two levels. We have described a flexible Bayesian latent variable model for the situation where several binary outcomes are measured to assess an intervention effect in a cluster-randomized clinical trial with 3-level hierarchical data.

Chapter 4 illustrates the single latent trait model in a cluster randomized clinical trial of three interventions to improve the processes of care for outpatients with pneumonia. Four binary outcomes are collected at the patient-level and clustered at the provider and clinic site levels. Simulation studies are conducted to check the algorithm and computational implementation. Chapter 5 extends the one latent trait model to a two-latent trait model using eight outcomes from both outpatient and inpatient care.

This latent modeling approach provides a comprehensive way to analyze multivariate hierarchical data. The method not only allows assessment of intervention effects with respect to multiple outcomes by borrowing strength across outcomes, but also assesses the relationship between outcomes, identifies those outcomes that carry the most information about the latent trait(s), and provides a summary measure of the "quality of care" at each clinical site. Although the frequently used simple raw average methodology could also compute summary scores, it cannot be directly applied to clusters with missing data, or incorporate covariance information or relationships between latent variables, in contrast to model based approach (Skrondal and Rabe-Hesketh, 2004).

The proposed single latent trait model makes one important assumption, that a univariate latent variable explains the observed pattern in the data and that the intervention affects the outcomes in the similar way (same direction) , because we standardized by fixing the variance of the latent trait as 1 and truncating the factor loading $\{\lambda_h\}$ below 0 for identification. This assumption is reasonable for the example of outpatient data used in Chapter 4 to illustrate the one latent trait model, because the individual outcome analyses show that the intervention has a positive effect on each outcome. Although *oxygen assessment* is poorly correlated with the other outcomes, indicating this outcome might measure a different underlying construct, removing it from the latent trait results in slightly bigger DIC, suggesting that the model with one trait constructed from the four outcomes fits better. In Chapter 5, model identification is achieved by "anchoring" by fixing the scale of factor loading instead of 'standardization' and truncation, which relaxes the assumption of a common intervention effect, but loses the interpretation that higher latent scores indicate higher quality of pneumonia care when the estimates of factor loadings have different signs across outcomes. However, in order to model the covariance

structure between multiple latent traits, we used ‘anchoring” instead of “standardization”, because “standardization” constraint the variance of latent trait to be 1. In Chapter 5, we refit one latent trait model using outpatient data by “anchoring”, the results are consistent with the results in Chapter 4 using “standardization”.

Due to the complexity of the model framework, in practice, latent variable models for hierarchical data could be framed in different ways for convenience. We frame the one latent trait model in Chapter 4 by specifying the within-site and between-site models separately, to better interpret the within-in site and between-site variation for cluster-randomized hierarchical clinical data in the EDCAP study. The model could be extended by adding additional covariates to the within-site model, the between-site model and/or the model of latent variable level, as appropriate. In Chapter 5, we frame the model using the recent developed very general model framework of Generalized Linear Latent and Mixed Models (GLLAMMs) (Skrondal & Rabe-Hesketh, 2007; Skrondal & Rabe-Hesketh, 2004), by writing two submodels: a response model and a structural model. The response model constructs latent variable(s) form the observed responses and the structural model regresses the latent variables as a function of intervention effect. Although the models were framed differently in Chapters 4 and 5, the underlying mathematical mechanisms are similar.

Our models make one important assumption of conditional independence of outcomes given the latent traits, which plays a central role in latent variable models. In further work, we will consider extending the model to relax this local independence assumption to allow conditional dependence between outcomes. Although in our work, we considered the situation where the multiple outcomes are binary, different outcome types, including both continuous and discrete endpoints can be incorporated into the proposed model framework. Another extension of

this hierarchical model is to relax the assumption of homogeneity of variance, to examine whether interventions influence the degree of heterogeneity (variation) across health care providers. In this situation, estimates of variance components are of as great an interest as the regression coefficients. This joint modeling has been done for a single outcome (Hedeker, Mermelstein, and Demirtas, 2008), but not for multiple outcomes using a latent construct.

Finally, our proposed two latent trait model models the outpatient and inpatient traits distinctly, which is reasonable for the EDCAP data because the two traits are poorly correlated conditional on the intervention effect. When the two traits are highly correlated, a possible extension would be to generate an overall trait constructed from those two individual latent traits, i.e., to estimate an overall intervention effect on this single latent trait or use this single trait to profile the health care providers.

In summary, this latent modeling approach provides a comprehensive way to analyze multivariate hierarchical data. The method not only allows assessment of intervention effects with respect to multiple outcomes, but also assesses the relationship between outcomes, identifies those outcomes that carry the most information about the latent trait(s), and provides a summary measure of the "quality of care" at each clinical site. Our work extends existing methods to model multivariate binary outcomes in a three-level hierarchical setting using one-latent trait and two-latent trait models, and assesses the relationships between multiple latent traits. A practical application demonstrates potential usefulness of this approach to quantify intervention effects with regard to multiple endpoints in a cluster-randomized clinical trial.

APPENDIX A

INDIVIDUAL OUTCOME DATA ANALYSIS

To explore the data, random logistic regression models accounting for clustering effect at site level and provider level were fitted for each outcome separately, to estimate the intervention effect adjusting for patient risk levels.

Both a Bayesian approach with WinBUGS and a maximum likelihood approach with the Stata Gllamm procedure were used to fit the models. Similar results were obtained based on the two different procedures. This demonstrates that the computation algorithm and the non-informative prior specification in our Bayesian approach are reasonable. Posterior summaries and maximum likelihood estimators of the intervention effect were shown in table 6-1. In total, 24 pairwise comparisons were conducted for the 8 outcomes. We noticed that the information derived from these individual outcome measures separately is overwhelming and not consistent across all outcomes. Furthermore, with this multidimensional outcome data, we expect some relationship among the outcomes, so that modeling outcomes independently would result in a loss of efficiency.

Table A-1 Bayesian and maximum likelihood estimations of intervention effects from individual outcome data analysis

Outcomes	comparisons	Posterior summaries				Maximum likelihood method		
		Mean	Median	sd	95% C.I.	Est.	S.E.	95% C.I.
Outpatient processes								
Oxygen assessment	Moderate vs. Low	0.11	0.12	0.70	(-1.27, 1.46)	0.10	0.61	(-1.09, 1.29)
	High vs. Low	0.42	0.43	0.74	(-1.05, 1.86)	0.40	0.63	(-0.82, 1.63)
	High vs. Moderate	0.31	0.32	0.62	(-0.95, 1.50)	0.30	0.53	(-0.73, 1.34)
First dose of antibiotics in ED	Moderate vs. Low	0.24	0.24	0.43	(-0.61, 1.08)	0.26	0.39	(-0.50, 1.01)
	High vs. Low	1.94	1.93	0.47	(1.01, 2.90)	1.90	0.42	(1.08, 2.72)
	High vs. Moderate	1.70	1.70	0.42	(0.88, 2.53)	1.65	0.37	(0.92, 2.37)
Treatment with compliant antibiotic therapy in ED	Moderate vs. Low	0.18	0.15	0.49	(-0.71, 1.22)	0.19	0.43	(-0.66, 1.04)
	High vs. Low	1.78	1.76	0.47	(0.90, 2.73)	1.78	0.44	(0.93, 2.64)
	High vs. Moderate	1.60	1.60	0.41	(0.81, 2.39)	1.59	0.38	(0.86, 2.33)
Compliant antibiotic therapy on discharge	Moderate vs. Low	0.75	0.74	0.40	(-0.04, 1.56)	0.77	0.35	(0.08, 1.47)
	High vs. Low	0.97	0.96	0.42	(0.12, 1.84)	0.99	0.36	(0.28, 1.70)
	High vs. Moderate	0.22	0.23	0.36	(-0.46, 0.94)	0.22	0.32	(-0.41, 0.85)
Inpatient processes								
Oxygen assessment	Moderate vs. Low	1.56	1.50	0.99	(-0.22, 3.67)	1.34	0.71	(-0.04, 2.73)
	High vs. Low	0.22	0.20	0.86	(-1.43, 2.00)	0.21	0.61	(-0.97, 1.40)
	High vs. Moderate	-1.34	-1.30	0.90	(-3.28, 0.37)	-1.12	0.66	(-2.43, 0.17)
Blood cultures before antibiotic administration	Moderate vs. Low	0.12	0.12	0.31	(-0.48, 0.74)	0.18	0.27	(-0.36, 0.72)
	High vs. Low	0.91	0.91	0.31	(0.31, 1.52)	0.94	0.28	(0.40, 1.48)
	High vs. Moderate	0.79	0.79	0.28	(0.22, 1.36)	0.76	0.25	(0.27, 1.25)
Antibiotic administration within 4 h	Moderate vs. Low	0.08	0.08	0.33	(-0.57, 0.75)	0.11	0.29	(-0.47. 0.68)
	High vs. Low	-0.13	-0.13	0.34	(-0.78, 0.55)	-0.11	0.29	(-0.68, 0.46)
	High vs. Moderate	-0.21	-0.21	0.29	(-0.79, 0.55)	-0.22	0.26	(-0.73, 0.30)
Treatment with compliant antibiotic therapy in ED	Moderate vs. Low	0.67	0.66	0.35	(0.03, 1.37)	0.62	0.34	(-0.04, 1.29)
	High vs. Low	1.30	1.30	0.35	(0.64, 2.01)	1.27	0.34	(0.61, 1.93)
	High vs. Moderate	0.64	0.64	0.33	(-0.02, 1.28)	0.65	0.30	(0.05, 1.24)

APPENDIX B

SELECTED WINBUGS FIGURES

Gibbs sampling history plots (trace plot) and posterior density plots are given for three parameters of the intervention effect. We examined convergence of the two Monte Carlo Markov chains by checking the trace plot. In Chapter 4 of one latent trait model, each parameter of interest becomes stationary by 3,000 iterations, indicating that the convergence has been reached by 3,000 iterations. In Chapter 5 of two-latent trait model, we used the first 10000 iterations as burn in. The posterior density plots for parameters show unimodal distributions, which are nearly symmetric, and look close to normal.

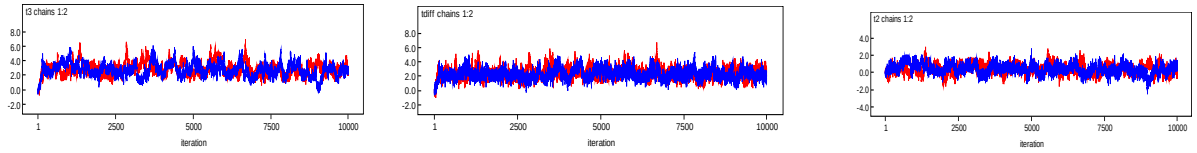


Figure B-6-1 Gibbs sampling trace plots of outpatient intervention effect of moderate vs. low (r_1), high vs. low (r_1), and high vs. moderate (r_2- r_1) sequentially, one latent trait model in Chapter 4

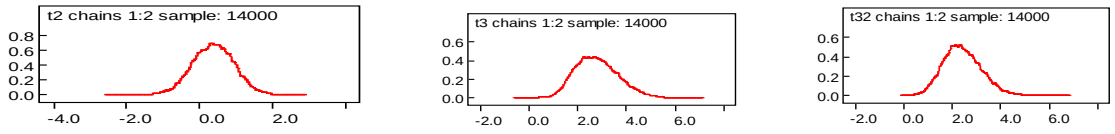


Figure B-6-2 Posterior density plots for parameters of outpatient intervention effect of moderate vs. low (r_1), high vs. low (r_1), and high vs. moderate (r_2- r_1) sequentially, one latent trait model in Chapter 4

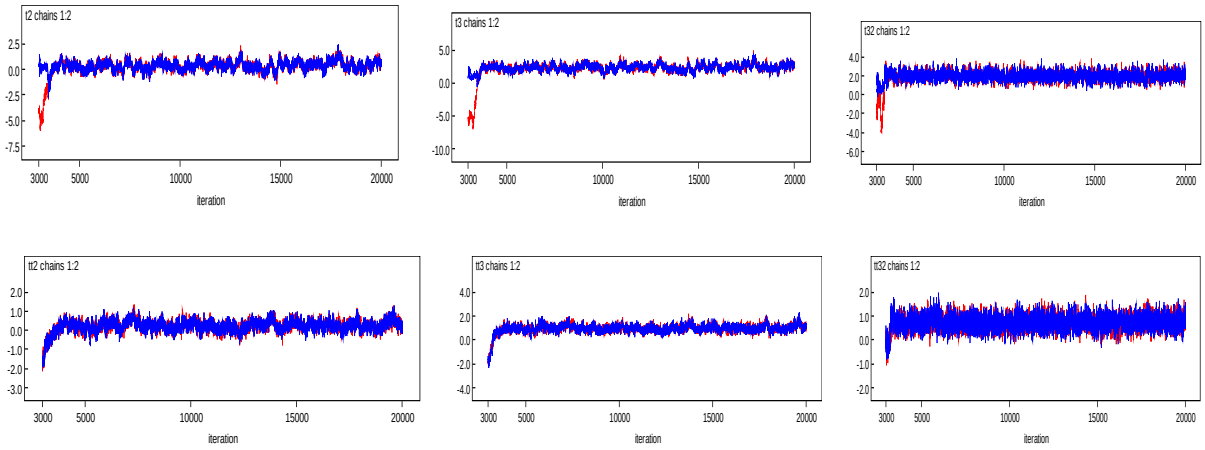


Figure B-6-3 Gibbs sampling trace plots of outpatient and inpatient intervention effect of moderate vs. low (r_1), high vs. low (r_1), and high vs. moderate (r_2- r_1) sequentially, one latent trait model in Chapter 5

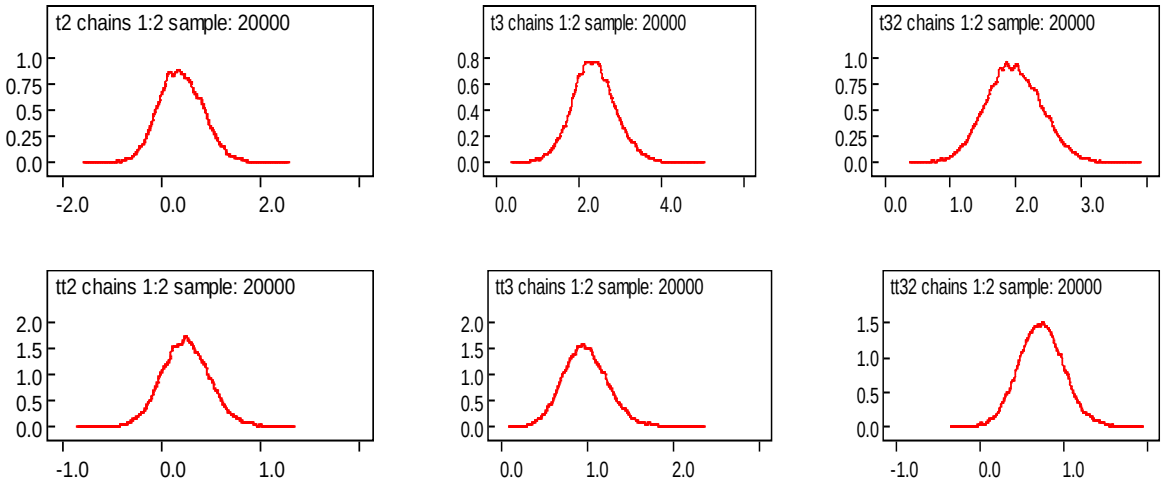


Figure B-6-4 Posterior density plots for parameters of outpatient and inpatient intervention effect of moderate vs. low (r_1), high vs. low (r_1), and high vs. moderate (r_2- r_1) sequentially, one latent trait model in Chapter 5

BIBLIOGRAPHY

Andrew Gelman (2006) Prior distribution for variance parameters in hierarchical models. Bayesian Analysis 1(3): 515-533.

Baker FB (1992) Item Response Theory: Parameter Estimation Techniques, New York: Marcel Dekker.

Bentler PM, Weeks DG (1980) Linear structural equations with latent variables. Psychometrika 45: 289-308.

Bentler PM, Stein JA (1992) Structural equationmodels in medical research. Statistical Methods in Medical Research 1: 159-181.

Blumenthal D (1996) Quality of care-what is it?-part one of six. The New England Journal of Medicine 335:891-894.

Bock RD, Aitkin M (1981) Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. Psychometrika, 46, 443-459.

Bollen KA. (1989) Structure equations with Latent Variables. Wiley, New York.

Bollen KA, Long JS (eds) (1993) Testing Structural Equation Models. Newbury Park, California: Sage.

Butler SM, Louis TA (1992) Random Effects Models with Non-Parametric Priors. Statistics in Medicine 11: 1981-2000.

Catalano PJ, Ryan LM (1992) Bivariate latent variable models for clustered discrete and continuous outcomes. Journal of the American Statistical Association 87: 651-658.

Choi K and Seltzer M (2010) Modeling Heterogeneity in Relationships Between Initial Status and Rates of Change: Treating Latent Variable Regression Coefficients as Random Coefficients in a Three-Level Hierarchical Model. Journal of Educational and Behavioral Statistics. 35: 1 54-91.

Croon M, Bolck A (1997) On the use of factor scores in structure equations models. Tech. Report 97.10.102/7 Work and Organization Research Center, Tilburg University

Daniels MJ, Normand S-LT (2006) Longitudinal profiling of health care units based on continuous and discrete patient outcomes. Biostatistics 7(1):1-15.

Donaldson GW (2003) General linear contrasts on latent variable means: structural equation hypothesis tests for multivariate clinical trials. Statistics in Medicine 22: 2893-2917.

Duan N, Manning WG, Morris CN, Newhouse JP (1983) A Comparison of Alternative Models for the Demand for Medical Care. Journal of Business & Economic Statistics 1(2): 115-126.

Dunson DB (2003) Bayesian latent variable models for clustered mixed outcomes. Journal of the Royal Statistical Society B 62: 355-366.

Eye AV, Clogg CC (eds) (1994) Latent Variable Modeling: Applications for Development Research, Thousand Oaks, Sage: Thousand Oaks, California.

Foulley JL, San Cristobal M, Gianola D, Im S (1992) Marginal likelihood and Bayesian approaches to the analysis of heterogeneous residual variances in mixed linear Gaussian models. Comput. Stat. Data Anal 13(3):291-305.

Gelman A, Carlin JB, Stern HS, Rubin DB (1995) Bayesian Data Analysis. Chapman & Hall: New York, U.S.A.

Gelman Andrew (2006) Prior distribution for variance parameters in hierarchical models. Bayesian Analysis 1(3): 515-533.

Geman S, Geman D (1984) Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. IEEE Transactions On Pattern Analysis and Machine Intelligence 6: 721-741.

Gilks WR, Thomas A, Spiegelhalter DJ (1994) A language and program for complex Bayesian modeling. The Statistician 43(1):169-177.

Gray SM, Brookmeyer R (1998) Estimating a treatment effect from multidimensional longitudinal data. Biometrics 54:976-988.

Hastings WK (1970) Monte Carlo sampling methods using Markov chain and their application. Biometrics 57: 97-109.

Hedeker D, Mermelstein RJ and Demirtas (2008) An Application of a Mixed-effects Location Scale Model for Analysis of Ecological Momentary Assessment (EMA) Data. Biometrics 64:627-634.

Jöreskog KG, Goldberger AS (1975) Estimation of a model with multiple indicators and multiple causes of a single latent variable. Journal of the American Statistical Association 70: 631–39.

Jöreskog KG (1977) Structural equation models in the social sciences. Specification, estimation and testing. In Krishnaiah. PR ed. Applications of Statistics. North-Holland 265–87.

Landrum MB, Bronskill SE, Normand SL (2000) Analytic methods for constructing cross-sectional profiles of health care providers. Health Services & Outcomes Research Methodology 1:23-48.

Landrum MB, Normand S-LT, Rosenheck RA (2003) Selection of related multivariate means: Monitoring psychiatric care in the Department of Veterans Affairs. Journal of the American Statistical Association 98(461):7-16.

Lee SY, Song XY (2003) Bayesian analysis of structural equation models with dichotomous variables. Statistics in Medicine 22: 3073-3088.

Lee SY, Song XY (2004) Bayesian model comparison of nonlinear structural equation models with missing continuous and ordinal data. British Journal of Mathematical and Statistical Psychology 57: 131-150.

Legler, JM, Lefkopoulu M, Ryan LM (1995) Efficiency and power of tests for multiple binary outcomes. Journal of the American Statistical Association 90:680-693.

Legler JM, Ryan LM (1997) Latent variable models for multiple birth outcomes. Journal of the American Statistical Association 92: 13-20.

Lin X, Ray J, Harlow SD (1997) Linear mixed models with heterogeneous within-cluster variance, Biometrics 53: 910-923.

Loannis Ntzoufras (2009) Bayesian Modeling Using Winbugs. John Wiley& Sons, Inc., Hoboken, New Jersey.

Lord FM (1952) A theory of test scores. Psychometric Monograph 7, Psychometric Society.

Louis TA (1982) Finding the observed information matrix when using the EM algorithm. Journal of the Royal Statistical Society, Series B 44:226-233.

Metropolis N et al. (1953) Equations of state calculations by fast computing machine. Journal of Chemical Physics 21: 1087-1091.

Miller RG (1981) Simultaneous Statistical Inference. New York: Springer-Verlag.

Morrison DF (1976) Multivariate Statistical Methods. New York: McGraw-Hill.

Muthén BO (1984) A general structural equation model with dichotomous, ordered categorical and continuous latent indicators. Psychometrika 49: 115–32.

O'Brien PC (1984) Procedures for comparing samples with multiple endpoints. Biometrics 40: 1079-1087.

Pocock SJ, Geller NL, Tsiatis A (1987) The analysis of multiple endpoints in clinical trials. Psychometrika 69: 167–190.

Rabe-hesketh S, Skrondal A, Pickles A (2004) Generalized multilevel structural equation modeling. Psychometrika, 69: 167-190.

Rabe-Hesketh S, Skrondal A, Pickles A (2005) Multilevel and structural equation modeling of continuous, categorical and event data. College Station, TX: Stat Press

Rabe-Hesketh S, Skrondal A (2008) Classical latent variable models for medical research. Statistical Methods in Medical Research 17: 5-32.

Sammel MD, Ryan LM (1996) Latent variable models with fixed effects. Biometrics 52 (92): 650-663.

Sammel MD, Ryan LM, Legler JM (1997) Latent variable models for mixed discrete and continuous outcomes. Journal of the Royal Statistical Society, Series B 59:667-678.

Sammel MD, Lin X, Ryan LM (1999) Multivariate linear mixed models for multiple outcomes. Statistics in Medicine 18: 2479-2492.

Schoenberg R, Richtand C (1984) An application of the EM method to the maximum likelihood estimation of multiple indicator and factor analysis models. Sociological Methods and Rresearch, 13, 127-150

Sik-Yum Lee (2007a) Structural Equation Modeling: A Bayesian Approach. New York: Wiley.

Sik-Yum Lee (2007 b) Handbook of Latent Variable and Related Models. North Holland.

Skrondal A, Rabe-Hesketh S (2004) Generalized Latent Variable Modeling: Multilevel Longitudinal, and Structural Equation Models. Chapman & Hall/CRC.

Skrondal A, Rabe-Hesketh S (2007) Latent Variable Modeling: A Survey. Scandinavian Journal of Statistics 34:712-745.

Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002) Bayesian measures of model complexity and fit. Journal of Royal Statistical Society, Series B (statistical Methodology) 64(4): 583-639.

Spiegelhalter DJ, Thomas A, Best NG, Lunn D (2003) WinBugs User Manual. Version 1.4 Cambridge, UK: MRC Biostatics Unit.

Tanner MA, and Wong WH (1987) The calculation of posterior distributions by data augmentation. Journal of the American Statistical Association 82: 528-540.

Teixeira-Pinto A, Normand S-L T (2008) Statistical methodology for classifying units on the basis of multiple related measures. Statistics in Medicine 27:1329-1350.

Verbeke G, Lesaffre E (1996) A Linear Mixed-Effects Model With Heterogeneity in the Random-Effects. Population 91 (433): 217-211.

Yealy DM, Auble TE, Stone RA, Lave JR, Meehan TP, Graff LG, Fine JM, Obrosky DS, Edick SM, Hough LG, Tuozzo K, Fine MJ (2004) The emergency department community acquired pneumonia trial: methodology of a quality improvement intervention. Annals of Emergency Medicine 43(6):770-782.

Yealy DM, Auble TE, Stone RA, Lave JR, Meehan TP, Graff LG, Fine JM, Obrosky DS, Mor MK, Whittle J, Fine MJ (2005) Effect of Increasing the Intensity of Implementing Pneumonia Guidelines. Annals of Internal Medicine 143(12):881-895.

Zhang Z, McArdle JJ, Wang L, Hamagami FA (2008) SAS interface for Bayesian analysis with WinBUGS. Structural Equation Modeling 15(4): 705-728.