Domain-Level Covariance Analysis for Survey Data with Structured Nonresponse

A. James O'Malley and Alan M. Zaslavsky[1]

Abstract

Health care quality surveys in the United States are administered to individual respondents (hospital patients, health plan members) to evaluate performance of health care units (hospitals, health plans). For better understanding and more parsimonious reporting of dimensions of quality, we analyze relationships among quality measures at the unit level. Rather than specifying a full parametric model for the observed responses and the non-response patterns at the lower (patient) level, we first fit generalized variance-covariance functions that take into account nonresponse patterns in the survey responses. We then specify a likelihood function for the unit mean responses using these generalized variance-covariance functions, letting us model directly the quantities we want to report. Because the response scales are bounded we assume that the unit means follow a truncated multivariate normal distribution. We calculate maximum likelihood estimates using the EM algorithm or drawing directly from Bayesian models using Markov-chain Monte Carlo. Finally factor analysis is performed on the between-unit covariance matrices obtained from the fitted models. Using posterior draws we assess posterior distributions of the number of selected factors and the assignment of items to groups under conventional rules. We compare maximum likelihood estimates of this factor structure to those from several Bayesian models with different prior distributions for the between-unit covariance. Results are presented using data from the Consumer Assessment of Healthcare Providers and Systems (CAHPS®) survey of Medicare Advantage health plans.

KEY WORDS: Bayesian, EM algorithm, Factor analysis, Hierarchical model, Markov-chain Monte Carlo, Covariance matrix priors.

[1]A. James O'Malley is Assistant Professor of Statistics (E-mail: *omalley@hcp.med.harvard.edu*) and Alan M. Zaslavsky is Professor (E-mail: *zaslavsky@hcp.med.harvard.edu*), Department of Health Care Policy, Harvard Medical School, Boston, MA 02115-5899. Research for this article was supported by grant U18 HS09205-06 from the U.S. Agency for Healthcare Quality Research and contract 500-95-007 from the U.S. Centers for Medicare and Medicaid Services.

1. Introduction

In the Consumer Assessments of Healthcare Providers and Systems (CAHPS®), surveys are conducted to elicit reports on experiences with health plans (entities administering health care) from enrolled members (Crofton, Lubalin, and Darby 1999; Schnaier et al. 1999; Goldstein et al. 2001). Such information is used by members, physicians, and payers to compare health plans and to facilitate quality improvement in health care. Optimal reporting of information about the quality of the care and services provided by health plans depends on identifying the important dimensions of quality and on reliably evaluating health plans' performance in each dimension. Recently this endeavor has been extended to assessment of other entities (hospitals, physician groups, nursing homes) through development of CAHPS surveys for patients with particular experiences, such as hospital stays (O'Malley and Zaslavsky 2005), or specific conditions (e.g., end-stage renal disease). These present similar analytical issues.

Factor analysis is commonly used to summarize patterns of variation among the items contained in the survey and thus identify the dimensions along which quality should be measured. The appropriate units of such analyses depend upon the questions of interest (Zaslavsky, Beaulieu, Landon, and Cleary 2000a). An individual-level analysis identifies items that are scored similarly by plan members. Correlations among items at the plan level indicate whether plans that perform well on one measure also perform well on others, and thus can guide summarization of the relative performance of different health plans.

The multilevel structure of CAHPS data provides challenges for estimation of the domain-level covariance matrix. (By "domain" we mean a unit of analysis with many elements, in this application a health plan or geographically defined portion of one in a given year.) If we were to ignore individual-level variation and analyze the raw means, correlated sampling error due to individual-level variation would confound plan-level relationships. Previous analyses used method-of-moments (MOM) estimators of between-domain covariance (Zaslavsky et al. 2000a; Zaslavsky and Cleary 2002). Although simple to implement, this approach lacks the flexibility to accommodate varying sample sizes across items and plans. Because domains receive equal weights, estimates are unstable when some domains have small sample sizes for some items. In the studies referenced above, domains were excluded from the analysis if the number of respondents to any item was below a certain threshold. This procedure is suboptimal because information is discarded arbitrarily. Furthermore, interval estimation and hypothesis testing can be tricky (Zaslavsky and Cleary 2002).

These limitations motivated us to fit multivariate hierarchical models to these data to estimate a domain-level covariance matrix. The challenge of fitting this model is compounded by two characteristics of CAHPS data. Firstly, CAHPS items have ordinal response formats with variable numbers of response options throughout the survey (ranging from 11-point to 2-point scales). Because the response scales are bounded sampling variances and covariances depend on the mean response. For example, the variance of the responses must be zero when the mean equals the minimum or maximum value. Scales with two points (e.g., yes/no) have the binomial variance function; with more points on the scale, the form of the variance function becomes more complex.

Secondly, many items are answered only by respondents who used particular services or had particular needs, as determined by screener items, creating pockets of structured nonresponse. This makes it difficult to model individual-level responses. Response rates to different CAHPS items vary tremendously, in this data set from 4% to 97% (O'Malley and Zaslavsky 2005). In the accepted approach to analysis of CAHPS data, the domain means for respondents are the relevant summary for each item, regardless of the varying response patterns, because these means reflect the assessments of those who used or needed the services in question.

To account for individual-level variation while avoiding direct modeling of the complicated response patterns we analyze the domain mean responses as in Fay and Herriot (1979), rather than directly modeling the individual-level data. The sampling variances and covariances of the domain means are expressed as functions of the domain means and other summary statistics such as the proportion of times a particular response pattern occurs. These generalized variance and covariance functions (GVCFs) were derived for CAHPS data by O'Malley and Zaslavsky (2005). GVCFs facilitate estimation of a multivariate hierarchical model by expressing the sampling covariance matrix for each domain as a function of estimates of the mean, thus avoiding reliance on direct estimation when such estimates are imprecise. Like Fay and Herriot we use the sample estimate of a variance or covariance of a domain mean when its expected error is smaller than that of the predictions from the GVCF.

In this study we implement both maximum likelihood and Bayesian approaches, focusing on the latter because it provides a flexible basis for inferences. Our application differs from those in which an underlying theory supports a covariance structure with a specific parsimonious form, as in spatial analysis or image processing. We face the additional challenge of modeling a covariance matrix with many unknown parameters. Because effects of the prior on inferences for unstructured covariance matrices are not well known, particularly in the context of hierarchical linear models, we consider and compare several prior distributions.

In the next section we describe the CAHPS data set that motivated this work. We introduce the hierarchical model for likelihood-based inference in Section 3.1 and describe the GVCFs for the sampling variation in Section 3.2. Section 3.3 develops a Bayesian model and proposes several priors for the between-domain covariance matrix, including priors that separately model the diagonal and non-diagonal components of the between-domain covariance matrix as well as the usual inverse-Wishart prior. Estimation methods and algorithms are described in Section 4, and results are presented in Section 5. The paper concludes with discussion in Section 6.

2. Medicare CAHPS Data

Since 1997 a CAHPS survey has been administered to beneficiaries of U.S. Medicare managed care plans, comprising over 1997-2001 a population of around 6 million, including persons over age 65 (about 93%) and persons with disabilities (about 7%). Our data set consists of all responses to the Medicare CAHPS survey during 1997-2001, encompassing 381 reporting units, each sampled in 1 to 5 years, for a total of 1195 reporting unit-year combinations (reporting domains) with 705,848 responses. Each reporting domain consists of the enrollees of a plan (or geographically defined portion of one) in a year, from whom equal-sized systematic samples are drawn. Because samples are drawn independently each year, patients may be sampled in multiple years. However, repeated sampling is rare and can be overlooked for our analysis.

Plan mean scores (perhaps after formation of composite scales) on the various survey items are calculated and reported to consumers. The median total number of respondents in a domain is 474 and 96% of domains have at least 200 respondents. Therefore, even if an item has a low response rate (e.g., less than 20%) we expect the distribution of the mean scores to be close to normal for most domains and items.

Each year the survey included around 31 report items that were selected to assess particular experiences of the patient, and 4 items that elicited general ratings of care. Report items use a 4-point ordinal "frequency" scale (never/sometimes/usually/always), or a 3-point ordinal "problem" scale (not a problem/somewhat a problem/a big problem), or are dichotomous (no/yes). Rating items use an 11-point (0–10) numerical scale. Complete item wordings and additional summary statistics appear in Zaslavsky et al. (2000a) and Zaslavsky and Cleary (2002).

Screener items determine which respondents did not use or need particular services or have particular needs, so they can be instructed to skip ahead in the survey. Because the values of skipped items are not meaningful, e.g., the convenience of getting durable medical equipment for a patient who reported having no need for such equipment, we do not impute missing values for ineligible respondents. Instead the plan's score for an item is defined as the mean response of those that appropriately responded. Almost all item nonresponse (over 98%) is due to properly skipped items.

3. Models

3.1. Hierarchical Model Structure

The relationship between the sample and population domain mean ratings for r items may be characterized by a two-stage multivariate hierarchical model. The vector of sample means of item responses for domain h ($h = 1, \ldots, n$) is denoted $\boldsymbol{Y}_h$, and assumed to satisfy:

$$\boldsymbol{Y}_h \dot{\sim} N(\boldsymbol{\theta}_h, \boldsymbol{V}_h), \tag{1}$$

indicating that $\boldsymbol{Y}_h$ has expectation $\boldsymbol{\theta}_h = E[\boldsymbol{Y}_h \mid \boldsymbol{\theta}_h]$ and covariance $\boldsymbol{V}_h = \text{var}[\boldsymbol{Y}_h \mid \boldsymbol{\theta}_h]$ with distribution "in the vicinity ($\dot{\sim}$)" of a Gaussian (Normal) density. Although $\boldsymbol{Y}_h$ could not actually be normally distributed due to boundary conditions, its distribution still is close enough to normal for most values of $\boldsymbol{\theta}_h$ for purposes of our inferences. In general, $\boldsymbol{Y}_h$ may be any summary statistic; the sample mean is, however, of the most interest to us. We assume that the sampling covariance of $\boldsymbol{Y}_h$ has the form

$$\boldsymbol{V}_h = V(\boldsymbol{\theta}_h, R_h, \boldsymbol{\alpha}, \boldsymbol{\beta}), \tag{2}$$

where R_h is the response pattern, and $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are unknown parameters. Let $\boldsymbol{C}_h$ denote the corresponding correlation matrix.

The second (structural) stage of the model describes the distribution of the mean ratings for each domain. Let $\mathcal{A}$ be the hyper-rectangle of the set of valid parameter values for $\boldsymbol{\theta}_h$ as determined by the numerical ranges of the corresponding survey items. The model for the domain means is

$$\boldsymbol{\theta}_h \dot{\sim} TN(\mathcal{A}; \boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{3}$$

where $TN(\mathcal{A}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the normal density truncated to the subspace $\mathcal{A}$. Because the mean and variance of $\boldsymbol{\theta}_h$ depend on $\mathcal{A}$ they do not equal their unrestricted counterparts $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$.

The model given by (1 – 3) follows Fay and Herriot (1979) in that we model aggregated data rather than single observations and express the second moment as a function of the first moment at the observation (e.g. patient) level. Modeling aggregate data avoids modeling patterns of missing data (from skips) at the individual level, which would greatly complicate the analysis for potentially little gain in precision of inferences at the domain level. The key assumptions of this approach are that the variances and covariances can be modeled adequately as functions of the mean (i.e., that no serious bias arises from model misspecification) and that the skip patterns are independent of domain characteristics. We believe that both of these assumptions are reasonable for CAHPS data; see O'Malley and Zaslavsky (2005) for more discussion.

3.2. Generalized Variance and Correlation Functions

Design-consistent variance estimators can generally be found using standard methods (e.g., Taylor series expansions). Under equal-probability sampling $\bar{\boldsymbol{Y}}_h$ is the design-consistent estimator of $\boldsymbol{\theta}_h$. However, design-based level I sampling variance and covariance estimates are unstable when the number of responses from a domain to some or all items is small. Also they do not reflect model-based inferences about population means that differ from sample means. We instead fit models that predict the sample variances and correlations from the domain means of the associated items. The material summarized in this section is presented more comprehensively in O'Malley and Zaslavsky (2005).

Let V_{hii} and V_{hij} denote the sampling variance and sampling covariance of the ith item and ijth pair of items in the hth domain respectively, and let n_{hi} denote the number of respondents to item i in domain h. The generalized variance function is assumed to have the form

$$\tilde{n}_{hi} V_{hii}^{\text{mod}} = V(\theta_{hi}) = \beta_{1i}\theta_{hi} + \beta_{2i}\theta_{hi}^2 \text{ for } i = 1, \ldots, r, \tag{4}$$

where $\tilde{n}_{hi}$ is a scaling constant incorporating the sample size and the design effect; $\tilde{n}_{hi} = n_{hi}$ under equal probability sampling. Because $\boldsymbol{V}_h$ must be positive definite, β_{1i} and β_{2i} are constrained to be positive. There is no problem if $n_{hi} = 0$ because $\boldsymbol{V}_h$ only enters the hierarchical modeling computations as the information matrix $\boldsymbol{V}_h^{-1}$, which is well defined when $n_{hi} = 0$.

We model the sample correlations because they are independent of the range of values, the response rates, and the sample sizes of the items. To account for the bounded scale of the correlations, we model the Fisher z-transform $z_{hij} = \log\{(1 + C_{hij})/(1 - C_{hij})\}$, where C_{hij} denotes the sample correlation between items i and j when jointly observed at domain h. The generalized correlation function we use (see O'Malley and Zaslavsky (2005) for justification) is given by

$$z_{hij}^{\text{mod}} = \alpha_{0ij} + \alpha_{1ij}\theta_{hi}\theta_{hj}. \tag{5}$$

The parameters $(\boldsymbol{\beta}, \boldsymbol{\alpha})$ of the GVCFs are estimated once (by regressing the sample variances and z-transformed correlations on the sample means) and then fixed for the remainder of the analysis. Because the GVCFs are not themselves of interest we treat $(\boldsymbol{\beta}, \boldsymbol{\alpha})$ as known constants and henceforth omit references to $(\boldsymbol{\beta}, \boldsymbol{\alpha})$. Although this simplification will result in underestimation of variability at the individual-level, it has much less impact on unit-level inferences especially when the unit sample sizes are large.

The model-based sampling covariance matrix for each domain is constructed using $\{\tilde{V}_{h11}^{\text{mod}}, \ldots, \tilde{V}_{hrr}^{\text{mod}}\}$ and $\tilde{\boldsymbol{z}}_h^{\text{mod}}$. Sampling covariances are calculated from the sample covariance of the ratings for the set of respondents who answered both items and the differences in mean responses for each item across the response states (responded or skipped) of the other item.

In general, shrinking direct estimates towards a model-based smoothed value improves precision (Fay and Herriot 1979, Ghosh and Rao 1994). Because model-based estimates are susceptible to lack of fit, for sufficiently large samples direct estimates are more accurate than model-based estimates. Furthermore, predicted correlations from our componentwise models are not guaranteed to form a positive definite matrix whereas the direct estimates are. Thus a better estimator might be a weighted average of model predictions and direct estimates that gives model-based estimates more weight for domains with small samples and direct estimates more weight for domains with large samples.

Let n_h denote the total number of individuals sampled in domain h, $\tilde{V}_{hii}^{\text{dir}}$ and $n_{hi}\tilde{V}_{hii}^{\text{pdir}} = (\sum_h n_{hi})^{-1} \sum_h n_{hi}\tilde{V}_{hii}^{\text{dir}}$ denote the associated direct and pooled direct variances respectively, and $\tilde{\boldsymbol{z}}_h^{\text{dir}}$, and $\tilde{\boldsymbol{z}}_h^{\text{pdir}}$ the associated correlation matrices. Because n_h is constant over items $\tilde{\boldsymbol{z}}_h^{\text{dir}}$ is a positive definite matrix if any of the direct correlation matrix estimates are positive definite. Then to make $\hat{V}_h$ positive definite regardless of $\boldsymbol{\theta}_h$ we compute:

$$\hat{V}_{hii} = t_h q_{hi} \tilde{V}_{hii}^{\text{mod}} + t_h(1 - q_{hi})\tilde{V}_{hii}^{\text{dir}} + (1 - t_h)\tilde{V}_{hii}^{\text{pdir}}, \text{ for } i = 1, \ldots, r, \text{ and} \tag{6}$$

$$\hat{\boldsymbol{z}}_h = t_h \bar{u}_h \tilde{\boldsymbol{z}}_h^{\text{mod}} + t_h(1 - \bar{u}_h)\tilde{\boldsymbol{z}}_h^{\text{dir}} + (1 - t_h)\tilde{\boldsymbol{z}}_h^{\text{pdir}}, \tag{7}$$

where q_{hi} denotes the ratio of the variance of the direct variance estimate to the sum of the variances of the direct and model-based variance estimates for item i in domain h, and $\bar{u}_h$ denotes the average of the analogous ratio over the elements of the associated correlation matrix.

Figure 1 illustrates how the three estimators are combined. If the correlation matrix formed from the model and direct estimates, $\bar{u}_h \hat{\boldsymbol{z}}_h^{\text{mod}} + (1 - \bar{u}_h)\hat{\boldsymbol{z}}_h^{\text{dir}}$, is positive definite we set $t_h = 1$. Otherwise we guarantee positive definiteness by setting $t_h \in [0, 1)$ such that the eigenvalues of the resulting covariance matrix are bounded away from 0 by $\epsilon > 0$. The sampling covariances are then determined from $\hat{V}_{hii}$, $\hat{V}_{hij}$, $\hat{\boldsymbol{z}}_{hij}$, and n_{hij} (the number of respondents to both items i and j in domain h) as in O'Malley and Zaslavsky (2005).

Because the weights $(\{q_{h1}, \ldots, q_{hr}\}, \bar{u}_h)$ rely solely on data within domains, they can be held constant throughout the level II computation. However, because t_h depends on $\boldsymbol{\theta}_h$ the relative weights of the estimators are subject to change while fitting the model. To simplify notation we do not explicitly refer to the model weights in the remainder of the article.

3.3. Prior Distributions for Bayesian Model

To complete the specifications of the Bayesian models we specify prior distributions for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. We assume that $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are a priori independent and that the prior for $\boldsymbol{\mu}$ is locally uniform, i.e., $p(\boldsymbol{\mu}) \propto 1$.

3.3.1. Inverse-Wishart Prior

We first consider the inverse-Wishart prior distribution for $\boldsymbol{\Sigma}$

$$\boldsymbol{\Sigma} \sim IW\left(b_0, \boldsymbol{B}_0\right), \tag{8}$$

where $b_0 \geq r$ is required for the density to be proper. The scalar degrees-of-freedom parameter $b_0 \geq r$ can be regarded as a prior sample size. An undesirable feature of the inverse-Wishart prior is that a single parameter controls the precision of all elements of $\boldsymbol{\Sigma}$.

3.3.2. Separation-Strategy Prior

A general way of enabling prior precision to vary across elements is to decompose $\boldsymbol{\Sigma}$ into components and specify separate priors for each component. Define $\boldsymbol{\Delta} = \text{diag}(\boldsymbol{\delta})$, where $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_r)$ and $\delta_i > 0$, let $\boldsymbol{\Phi}$ be a positive definite matrix, and let $\boldsymbol{\Sigma} = \boldsymbol{\Delta\Phi\Delta}$. We do not constrain $\boldsymbol{\Phi}$ to be a correlation matrix and so the model is over-parameterized. The associated correlation matrix is $\boldsymbol{\Omega} = \boldsymbol{\Omega}(\boldsymbol{\Phi}) = \text{diag}(\boldsymbol{\Phi})^{-1/2}\boldsymbol{\Phi}\text{diag}(\boldsymbol{\Phi})^{-1/2}$, where $\text{diag}(\boldsymbol{\Phi})$ retains the diagonal elements of Φ but has zeros elsewhere. We assume that $\log(\delta_i) \overset{\text{iid}}{\sim} N(s_0, v_0^2)$ for $i = 1, \ldots, r$ and consider two different families of priors for $\boldsymbol{\Phi}$ based on the separation strategy of Barnard, McCulloch, and Meng (2000) (henceforth, BMM).

We first consider an $IW(b_0, \boldsymbol{I})$ prior for $\boldsymbol{\Phi}$ where $\boldsymbol{I}$ denotes the identity matrix. From BMM the implied marginal distribution of $\boldsymbol{\Omega}$ is given by

$$f(\boldsymbol{\Omega}) \propto |\boldsymbol{\Omega}|^{-(b_0+r+1)/2} \left(\prod_i \omega^{ii}\right)^{-b_0/2} \tag{9}$$

where ω^{ii} is the ith diagonal element of $\boldsymbol{\Omega}^{-1}$. The marginal distribution of an individual non-diagonal element of $\boldsymbol{\Omega}$, ω_{ij} for $i \neq j$, is proportional to $(1 - \omega_{ij}^2)^{(b_0-r-1)/2}$, the density of a $\text{Beta}((b_0 - r + 1)/2, (b_0 - r + 1)/2)$ distribution. Therefore, setting $b_0 = r + 1$ yields a prior under which the marginal distribution of each correlation is uniform on $[-1, 1]$. Accordingly this is the "marginal uniform" or flat prior, which we denote $\text{SepStrat}_{\text{flat}}$. We also consider the cases $b_0 \in [r, r+1)$ and $b_0 > r+1$. As these priors induce marginal distributions of the individual correlations that are respectively bimodal (convex) and unimodal (concave), we call these $\text{SepStrat}_{\text{vex}}$ and $\text{SepStrat}_{\text{cave}}$ respectively. $\text{SepStrat}_{\text{vex}}$ has higher density for correlations far from 0 whereas $\text{SepStrat}_{\text{cave}}$ has higher density for correlations close to 0.

When $b_0 \geq r$ the posterior distribution associated with (9) is assured of being well defined because the prior is proper. Although improper priors with $b_0 < r$ such as $f(\mathbf{\Phi}) \propto 1$ and the Jeffreys prior $f(\mathbf{\Phi}) \propto |\mathbf{\Phi}|^{-(r+1)/2}$ may be used in non-hierarchical models, they yield improper posteriors in this model. Indeed $TN(\mathcal{A}; \boldsymbol{\mu}, t\mathbf{\Phi}_0)$ approaches the uniform on the bounded space $\mathcal{A}$ as $t \to \infty$, so the marginal likelihood of $\mathbf{\Phi}$ approaches a constant. However, the jointly uniform prior $f(\mathbf{\Phi}) \propto 1$ could be used because the space of correlation matrices is bounded, so the prior is itself proper (BMM).

If we used a truncated-normal rather than a log-normal distribution for δ_i, the separation strategy family of priors would be a multivariate version of Gelman's folded half-T prior (Gelman 2005). Although other priors may be used for δ_i, such as the truncated normal, we favor the log-normal because it has an unrestricted scale.

3.3.3. Incorporation of factor structure into prior

The SepStrat prior may be extended by incorporating information about the hypothesized factor structure, specifically about the number of underlying factors. We define the prior by assuming the density is proportional to the product of the $IW(b_0, \boldsymbol{I})$ density and a Dirichlet density for the ordered eigenvalues $(\lambda_1, \ldots, \lambda_r)$ of $\mathbf{\Omega}$, implying

$$f(\mathbf{\Phi}) \propto |\mathbf{\Phi}|^{-(b_0+r+1)/2} \exp\{-\text{tr}(\mathbf{\Phi}^{-1})/2\}. \prod_{k=1}^{r} \lambda_k^{(\kappa_k - 1)}, \tag{10}$$

where $\kappa_k \geq 0$. If a priori the dimension of the underlying factor structure is believed to be K one might make the first K elements of κ large and the remaining elements small.

4. Estimation

4.1. Bayesian Computation

To fit the Bayesian models we use Markov chain Monte Carlo (MCMC) methods with independent updates for $\theta_1, \ldots, \theta_n$, $\boldsymbol{\mu}$, and $\mathbf{\Sigma}$. Gibbs steps are used to update the parameters for which conditional posterior distributions are available; otherwise Metropolis-Hastings steps (Metropolis et al. 1953;

Hastings 1970) are used.

4.1.1. Augmented data model

Rather than drawing $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ directly under the truncated normal model $\boldsymbol{\theta}_h \sim TN(\mathcal{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ we augment $\{\boldsymbol{\theta}_h\}_{h=1:n} \sim TN(\mathcal{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ with enough data so that the combined data may be regarded as $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, allowing us to estimate $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ as the parameters of a multivariate normal distribution. Let N denote the size of the total sample (including the augmented data).

Given N draws from the unconstrained $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution, $n \sim \text{Bin}(N, \pi(\mathcal{A}))$ where $\pi(\mathcal{A}) = \text{pr}(\boldsymbol{\theta}_h \in \mathcal{A} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$. We specify the single observation unbiased prior (SOUP) $f(N) \propto 1/N$ (Meng and Zaslavsky 2002); this has the appealing property that $E[N \mid n]$ is an unbiased estimator of N.

The augmented model for the ratings is

$$\begin{aligned} &\text{Level I:} && \boldsymbol{Y}_h \mid \boldsymbol{\theta}_h \sim N(\boldsymbol{\theta}_h, V(\boldsymbol{\theta}_h)) \text{ for } h = 1, \ldots, n \\ &\text{Level II:} && \boldsymbol{\theta}_h \mid \boldsymbol{\mu}, \boldsymbol{\Sigma}, N \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \text{ for } h = 1, \ldots, N, \\ &\text{Level III:} && f(N) \propto N^{-1},\ f(\boldsymbol{\mu}) \propto 1,\ \text{and } \boldsymbol{\Sigma} \sim f(\boldsymbol{\Sigma}). \end{aligned}$$

The above model implies that the posterior distribution of $(N-n)|n, \boldsymbol{\mu}, \boldsymbol{\Sigma}$ is the negative binomial $\text{NB}(n, \pi(\mathcal{A}))$ distribution. Bayesian inferences based on this augmented model, obtained by integrating over N and $\{\boldsymbol{\theta}_h\}_{h=n+1,\ldots,N}$, are equivalent to those under the original model (Gelman, Carlin, Stern, and Rubin 1995, page 193).

The covariance of $TN(\mathcal{A}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$, denoted $\boldsymbol{\Sigma}_{\mathcal{A}}$, is estimated numerically from a large sample of draws from this distribution. The associated correlation matrix is denoted $\boldsymbol{\Omega}_{\mathcal{A}}$.

4.1.2. Conditional posterior of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$

Because $\boldsymbol{\mu} \mid N, \boldsymbol{\theta}, \boldsymbol{\Sigma} \sim N(\sum_{h=1}^{N} \boldsymbol{\theta}_h / N, \boldsymbol{\Sigma}/N)$, a Gibbs step is used to update $\boldsymbol{\mu}$.

Under the inverse-Wishart prior for $\boldsymbol{\Sigma}$ the conditional posterior for $\boldsymbol{\Sigma}$ is the inverse-Wishart distribution given by

$$\boldsymbol{\Sigma} \mid N, \boldsymbol{\theta}, \boldsymbol{\mu} \sim IW\left\{ b_0 + N, \boldsymbol{B}_0 + \sum_{h=1}^{N} (\boldsymbol{\theta}_h - \boldsymbol{\mu})(\boldsymbol{\theta}_h - \boldsymbol{\mu})^T \right\}.$$

A Gibbs step is used to draw samples of $\boldsymbol{\Sigma}$ in this case.

Under the other specifications the prior for $\boldsymbol{\Sigma}$ is implied by the priors for the scale parameter $\boldsymbol{\delta}$ and the matrix parameter $\boldsymbol{\Phi}$. The conditional posterior distribution of $\boldsymbol{\Phi}$ under the inverse-Wishart prior for $\boldsymbol{\Phi}$ in (9) is

$$\boldsymbol{\Phi} \mid N, \boldsymbol{\theta}, \boldsymbol{\mu} \sim IW\left\{ b_0 + N, I + \boldsymbol{\Delta}^{-1} \left(\sum_{h=1}^{N} (\boldsymbol{\theta}_h - \boldsymbol{\mu})(\boldsymbol{\theta}_h - \boldsymbol{\mu})^T \right) \boldsymbol{\Delta}^{-1} \right\},$$

where $\boldsymbol{\Delta} = \text{diag}(\delta_1, \ldots, \delta_r)$. Calculation of $\boldsymbol{\Omega} = \text{diag}(\boldsymbol{\Phi})^{-1/2}\boldsymbol{\Phi}\text{diag}(\boldsymbol{\Phi})^{-1/2}$ requires $\boldsymbol{\Phi}$ alone, while calculation of $\boldsymbol{\Sigma} = \boldsymbol{\Delta}\boldsymbol{\Phi}\boldsymbol{\Delta}$ also requires $\boldsymbol{\delta}$.

When $\log(\delta_i) \sim N(s_0, v_0^2)$ the log conditional posterior distribution of δ_i reduces to

$$\log f(\delta_i) = -(N+1)\log(\delta_i) - [\boldsymbol{\Phi}^{-1}]_{ii}V_{ii}/(2\sigma_i^2) - \frac{1}{\delta_i}\sum_{j\neq i}[\boldsymbol{\Phi}^{-1}]_{ij}V_{ij}/\delta_j - (\log(\delta_i) - s_0)^2/(2v_0^2) + \text{constant}, \tag{11}$$

where $[\boldsymbol{\Phi}^{-1}]_{ij}$ is the ijth element of $\boldsymbol{\Phi}^{-1}$ and V_{ij} is the ijth element of $\sum_h(\boldsymbol{\theta}_h - \boldsymbol{\mu})(\boldsymbol{\theta}_h - \boldsymbol{\mu})^T$. We use a Metropolis-Hastings step with a log t proposal distribution with 3 degrees of freedom to update each component of $\boldsymbol{\delta}$ in turn. The location parameter of the proposal density is the current value of δ_h and the scale parameter is adaptively chosen to make the acceptance rate roughly 0.44, as suggested by Gelman, Roberts, and Gilks (1996).

With the family of priors for $\boldsymbol{\Sigma}$ defined by (10) we take a Metropolis-Hastings step using as the proposal density the posterior under the $IW(b_0, \boldsymbol{I})$ prior. Because the contributions from the likelihood function and the priors on $\{\boldsymbol{\theta}_h\}$, $\boldsymbol{\mu}$, and $\boldsymbol{\delta}$ cancel, the acceptance probability depends only on the ratio of the Dirichlet densities for the eigenvalues of $\boldsymbol{\Phi}$. Therefore, at iteration t the acceptance probability is given by $\min\left\{1, \prod_{i=1}^r(\lambda_i^{(t)}/\lambda_i^{(t-1)})^{\omega-1}\right\}$.

4.1.3. Conditional posterior of $\boldsymbol{\theta}_h$ for $h = 1, \ldots, n$

The dependence of $\boldsymbol{V}_h$ on $\boldsymbol{\theta}_h$ prevents derivation of the posterior distribution of $\boldsymbol{\theta}_h$ in closed form. However, the Metropolis-Hastings algorithm only requires the posterior to be known up to proportionality. The log-conditional posterior of $\boldsymbol{\theta}_h$ is

$$\log l(\boldsymbol{\theta}_h) = \log|V(\boldsymbol{\theta}_h)|/2 - (\boldsymbol{Y}_h - \boldsymbol{\theta}_h)'V(\boldsymbol{\theta}_h)^{-1}(\boldsymbol{Y}_h - \boldsymbol{\theta}_h)/2 - (\boldsymbol{\theta}_h - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_h - \boldsymbol{\mu})/2 + \text{constant}.$$

If $n_{hi} = 0$ we set $[V(\boldsymbol{\theta}_h)^{-1}]_{ij} = [V(\boldsymbol{\theta}_h)^{-1}]_{ji} = 0$ for $j \in 1, \ldots, n$, so the likelihood function is well defined.

A multivariate-T proposal density with 3 degrees of freedom centered at the current value of $\boldsymbol{\theta}_h$ and truncated (by rejection) to $\mathcal{A}$ is used to update $\boldsymbol{\theta}_h$. The multivariate-T covariance matrix parameter is set to a scalar multiple of $(\boldsymbol{\Sigma}^{-1} + \boldsymbol{V}_h^{-1})^{-1}$, the posterior covariance for the conjugate normal model. The scale is chosen to obtain an acceptance rate near 0.23, as suggested by Gelman, Roberts, and Gilks (1996).

4.1.4. Conditional posterior of $\boldsymbol{\theta}_h$ for $h = n+1, \ldots, N$ and of N

Because the augmented $\boldsymbol{\theta}_h$ are not associated with observed data, their conditional posterior distributions are the prior, i.e., $\boldsymbol{\theta}_h \sim TN(\mathcal{A}^c; \boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\mathcal{A}^c$ is the complement of $\mathcal{A}$, for $h = n+1, \ldots, N$.

We simultaneously update N and $\{\boldsymbol{\theta}_h\}_{h=(n+1):N}$ by repeatedly drawing $\boldsymbol{\theta}_h \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ until exactly n values of $\boldsymbol{\theta}_h \in \mathcal{A}$ have been obtained, thus obtaining draws from $\text{NB}(n, \pi(\mathcal{A}))$. The total number of draws is the new value of N, the truncated values $\boldsymbol{\theta}_h \in \mathcal{A}^c$ update $\{\boldsymbol{\theta}_h\}_{h=(n+1):N}$, and the remaining values are discarded. This avoids explicit evaluation of $\pi(\mathcal{A})$.

4.2. Maximum Likelihood Estimation

To obtain maximum-likelihood estimates we use a Monte Carlo EM algorithm (Wei and Tanner 1990). To account for the truncation of $\boldsymbol{\theta}_h$ we assume that $(N-n)$ has a $\text{NB}(n, \pi(\mathcal{A}))$ distribution, as suggested by Dempster, Laird, and Rubin (1977) and McLachlan and Krishnan (1997, page 75). The E-step of the EM algorithm relies upon the conditional posterior distribution of $\boldsymbol{\theta}_h$ given $\boldsymbol{Y}_h$ (if observed) and $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. For each iteration the E-step involves running the same MCMC chain used for the Bayesian computations but with $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ fixed at their current estimates (Goggins, Finkelstein, Schoenfeld, and Zaslavsky 1998). The M-step updates $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ using the Monte Carlo estimates of the conditional posterior means of their sufficient statistics (functions of $\boldsymbol{\theta}$) obtained from the E-step.

The E-step evaluates the conditional posterior expectation of the sufficient statistics of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ given $\{\boldsymbol{\theta}_h\}_{h=1:N}$ and N. The sufficient statistics are $U(\boldsymbol{\theta}) = N^{-1}\sum_{h=1}^{N}(\boldsymbol{\theta}_h, \boldsymbol{\theta}_h\boldsymbol{\theta}_h^T)$. Let $\boldsymbol{\theta}_h^{(s,t)}$ and $N^{(s,t)}$ denote draw t of $\boldsymbol{\theta}_h$ and N obtained from the MCMC simulation used to evaluate the E-step at iteration s of the EM-algorithm. The E-step evaluates

$$E[U(\boldsymbol{\theta}) \mid \boldsymbol{Y}, \boldsymbol{\mu}^{(s)}, \boldsymbol{\Sigma}^{(s)}] = T^{-1}\sum_{t=1}^{T}(N^{s,t})^{-1}\sum_{h}\left(\boldsymbol{\theta}_h^{(s,t)}, \boldsymbol{\theta}_h^{(s,t)}\boldsymbol{\theta}_h^{(s,t)T}\right). \tag{12}$$

The M-step updates $\boldsymbol{\mu}^{(s)}$ and $\boldsymbol{\Sigma}^{(s)}$ by solving the likelihood equations for $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ at the values of the sufficient statistics computed in (12) to obtain

$$\boldsymbol{\mu}^{(s+1)} = T^{-1}\sum_{t=1}^{T}(N^{s,t})^{-1}\sum_{h}\boldsymbol{\theta}_h^{(s,t)}, \text{ and} \tag{13}$$

$$\boldsymbol{\Sigma}^{(s+1)} = T^{-1}\sum_{t=1}^{T}(N^{s,t})^{-1}\sum_{h}\boldsymbol{\theta}_h^{(s,t)}\boldsymbol{\theta}_h^{(s,t)T} - \boldsymbol{\mu}^{(s+1)}\boldsymbol{\mu}^{(s+1)T}. \tag{14}$$

After the algorithm has converged $\boldsymbol{\Sigma}_{\mathcal{A}}$ is estimated by the final value obtained for the covariance of the truncated distribution (i.e., computed using the final estimates of the non-truncated domain means as in the Bayesian computation).

4.3. Factor Analysis

To investigate the structure of $\boldsymbol{\Sigma}_{\mathcal{A}}$, we factor-analyze the estimated or sampled values of the associated correlation matrix $\boldsymbol{\Omega}_{\mathcal{A}}$. Factor analyses are conducted using the principle factor method with squared multiple correlations as initial communality estimates. In all analyses eigenvalues and loadings are reported for the reduced correlation matrix obtained by subtracting the initial estimates of the specific variances; thus the sum of the eigenvalues is less than the number of items. We use item-factor correlations to assess the relationship of each item to each factor. Analyses are based on the orthogonal decomposition unless the objective is to draw a substantive interpretation of the factor structure, in which case we use a non-orthogonal decomposition (the Promax criterion with Kaiser normalization) to rotate the factors so that the variability of the loadings within a factor is maximized. In the nonorthogonal rotation we use standardized regression coefficients to assess the relationship of each item to each factor.

Exploratory factor analysis describes the correlation structure and is not inherently inferential. However, the posterior distribution of $\boldsymbol{\Omega}_A$ in our Bayesian analyses, represented by samples, induce distributions of descriptive parameters such as the number of eigenvalues above a certain threshold (a conventional rule for determining the number of underlying factors) or the assignment of items to factors under the maximum loading rule. A novel feature of this analysis is the use of Bayesian analysis to evaluate the variability in the results obtained from an exploratory factor analysis; the usual practice is to apply factor analysis to a point estimate of the covariance or correlation matrix.

Because the ordering and signs of the columns of the factor loading matrix are arbitrary, we order the columns consistently prior to summarizing the factor analysis. Let $\boldsymbol{U}^{(t)} = (U_{ij}^{(t)})_{i=1,\ldots,r,j=1,\ldots,Q}$, where Q is the number of factors, denote the factor-loading matrix drawn at iteration t of T draws. First set $\boldsymbol{U}^{(\text{base})} = \boldsymbol{U}^{(1)}$. Then for $t = 2, \ldots, T$: (1) Compute the correlations between each pair of columns from $\boldsymbol{U}^{(t)}$ and $\boldsymbol{U}^{(\text{base})}$ and match up the columns of $\boldsymbol{U}^{(t)}$ and $\boldsymbol{U}^{(\text{base})}$ to maximize the sum of the absolute values of the correlations of corresponding columns; (2) Correspondingly, re-order the columns of loadings and the elements of $\boldsymbol{\lambda}^{(t)}$, reversing the sign of the column of $\boldsymbol{U}^{(t)}$ if the correlation is negative; (3) Update $\boldsymbol{U}^{(\text{base})}$ with $((t-1)\boldsymbol{U}^{(\text{base})} + \boldsymbol{U}^{(t)})/t$.

In step 2, usually each column of $\boldsymbol{U}^{(t)}$ had the highest correlation with a distinct column of $\boldsymbol{U}^{(\text{base})}$. At worst there were a few columns for which the optimal pairing required a search.

We summarize the factor loadings by calculating averages of various quantities across the draws. For example, to compute the posterior probability that item i loads highest on the kth factor, we compute the Monte Carlo average $T^{-1}\sum_{t=1}^{T} I\left(|U_{ik}^{(t)}| = \max_j\{|U_{ij}^{(t)}|\right)$, where $I(\text{event}) = 1$ if event is true and 0 otherwise. Let Q denote the number of eigenvalues of the reduced correlation matrix $\boldsymbol{\Omega}_A^{\text{red}} = \boldsymbol{\Omega}_A - \boldsymbol{\Psi}$, where $\boldsymbol{\Psi}$ contains the initial estimates of the specific variances of the factors on the diagonal and zeros elsewhere, that exceed its mean eigenvalue. The posterior probability that

$Q = k$ (i.e., that k factors are determined by the Guttman (1954) criterion) is

$$\text{pr}(Q \mid \boldsymbol{Y}) = T^{-1} \sum_{t=1}^{T} I \left(\sum_{i=1}^{r} I(\tilde{\lambda}_i^{(t)} > \bar{\tilde{\boldsymbol{\lambda}}}^{(t)}) = k \right),$$

where $\tilde{\lambda}_i$ denotes the ith eigenvalue of $\boldsymbol{\Omega}_A^{\text{red}}$. There is no such inference for maximum likelihood estimation because only a point estimate of $\boldsymbol{\Omega}_A$ is available.

4.4. Comparison of Results Across Approaches

We developed four measures to compare the factor structures obtained by maximum likelihood and the several Bayesian models. The first three measures assess the decisiveness and stability of the factor loadings within a single model, while the fourth measures the distance between the factor loadings for a pair of models.

To gauge the certainty of factor assignments we estimate the posterior probability an item is assignment to the factor with the highest posterior mean loading for that item. The "certainty" is estimated by

$$\text{Certainty}_i = T^{-1} \sum_{t=1}^{T} I \left(\max_k |U_{ik}^{(t)}| = k^* \right),$$

where $k^* : \left(\sum_{t=1}^{T} |U_{ik^*}^{(t)}| \right) = \max_k \left(\sum_{t=1}^{T} |U_{ik}^{(t)}| \right)$ is the factor with the highest posterior mean loading for item i. Certainty is a measure of the variation in the factor assignment of an individual item. The "Certainty score" is the average of Certainty_i over the items $1, \ldots, r$.

We define the "instability" of factor loadings under a model to be the average over items and factors of the posterior variances of the factor loadings, again averaged across the draws of U:

$$W = (rQ(T-1))^{-1} \sum_{i,k,t} (U_{ik}^{(t)} - \bar{U}_{ik})^2,$$

where $\bar{\boldsymbol{U}} = T^{-1} \sum_t \boldsymbol{U}^{(t)}$ is the mean factor loading matrix. Although $0 \le W \le 1$, W approaches 1 only if the columns of $\boldsymbol{U}_{t=1,\ldots,T}^{(t)}$ are not consistently ordered. Larger values of W indicate greater variability across the draws.

We measure how well items separate into groups by a scaled version of the Varimax criterion. Letting $\tilde{U}_{ik}^{(t)}$ denote the scaled loading obtained by dividing $U_{ik}^{(t)}$ by the square root of the tth draw of the communality for item i, the measure is given by

$$\text{Separation} = \frac{100Q}{Tr^2} \sum_{t,k} \left[\sum_i \tilde{U}_{ik}^{(t)4} - \frac{1}{r} \left(\sum_i \tilde{U}_{ik}^{(t)2} \right)^2 \right],$$

where the constant $100Q/r$ scales the Varimax criterion to a 0 to 100 scale. Separation is therefore proportional to the Monte Carlo average of the sum of the variances of the squared scaled loadings for each factor. The larger the Separation the more spread out the loadings are within the factors and thus the clearer the factor structure.

To evaluate the discordance of the item-factor loadings between two Bayesian models we compute

$$\text{Discordance} = 100B_{12}/((W_1 + W_2)/2 + B_{12}), \text{ where}$$

$$B_{12} = (rQ)^{-1} \sum_{i,k} (\bar{U}_{1,ik} - \bar{U}_{2,ik})^2,$$

W_1 and W_2 are the instability scores and $\bar{\boldsymbol{U}}_1$ and $\bar{\boldsymbol{U}}_2$ are the mean factor loading matrices, for each model. Discordance is thus the scaled ratio of the distance between the mean loading matrices and the total variation of the factor loadings; larger values imply more disagreement.

Although the above measures are designed to be evaluated on the draws from the Bayesian posterior distributions, they can also be applied to the maximum likelihood estimate by treating it as a single draw.

5. ANALYSIS OF CAHPS DATA SET

We compare estimates and other inferences from a näive analysis, in which the we ignore the hierarchical structure of the data and take the sample means to be the true domain means, to the maximum likelihood (ML) and four Bayesian approaches using the Medicare CAHPS 1997-2001 data ($n = 1195$, $r = 35$). The domains are health plans (or geographically defined portions of one) in a given year.

For the inverse-Wishart prior we set $b_0 = 10 + r + 1$ and $\boldsymbol{B}_0 = 0.011\boldsymbol{I} + 0.005\boldsymbol{J}\boldsymbol{J}'$, where $\boldsymbol{J}$ is a vector of r ones. Because the inverse-Wishart distribution is proper for $b_0 \geq r$ the proprietory of the posterior is assured while b_0 is small enough that the data are expected to dominate the analysis. The prior implies $E[\boldsymbol{\Sigma}] = \boldsymbol{B}_0/10$ which is consistent with a belief that the magnitudes of the between-domain variances and covariances are small. The value of $\boldsymbol{B}_0$ was chosen so that the ratio of the diagonal elements to the off-diagonal elements (around 0.45) approximated the correlation between CAHPS items often seen in practice.

By definition $b_0 = r + 1$ for $\text{SepStrat}_{\text{flat}}$. We set $b_0 = r$ for $\text{SepStrat}_{\text{vex}}$ and $b_0 = 3r$ for $\text{SepStrat}_{\text{cave}}$. The eigenvalue prior ($\text{SepStrat}_{\text{eig}}$) was defined by multiplying $\text{SepStrat}_{\text{flat}}$ with a Dirichlet density having 0.8 as the parameter for the first 4 (ordered) eigenvalues and 0.2 as the parameter for the remaining eigenvalues. In all Bayesian analyses an improper flat prior was assumed for $\boldsymbol{\mu}$, and for analyses involving the SepStrat prior independent log-normal priors with mean log(0.12) and variance 5 were assumed for $\{\sigma_i\}_{i=1,\dots,r}$.

5.1. Convergence Diagnostics

Because the E-step includes MCMC simulation the monotone convergence of the likelihood in the EM algorithm is lost (unless a huge number of values are drawn within each E-step), making it difficult to assess the convergence of the algorithm. We ran the EM algorithm using different starting points and monitored the convergence by plotting $(\boldsymbol{\mu}^{(s)}, \boldsymbol{\Sigma}^{(s)})$ against s. We terminated the algorithm when successive values had small random deviations about a common value. We set $T = 5s$ so that the E-step was evaluated with more precision the closer the algorithm was to convergence. To reduce the error from simulation the final estimates were averaged across the runs. We found that five parallel runs of 100 iterations each enabled estimates to be computed to a high degree of accuracy.

Convergence of the Markov chains used for Bayesian analyses was examined using CODA (Convergence Diagnostics and Output Delivery) software (Best, Cowles, and Vines 1995). The behavior of the chain was monitored using trace plots (Hellmich, Abrams, Jones, and Lambert 1998) of the sequence of draws for each parameter for different starting values and random number seeds. The corrected scale reduction factor of Brooks and Gelman (1998) indicated that a burn-in time of 5,000 iterations was appropriate (the 0.975 quantiles were generally less than 1.2).

When the half-width test of Heidelberger and Welch (1983) was applied with a test accuracy of 0.1 to a chain run for 50,000 iterations past burn-in it passed for all parameters, implying 50,000 was a sufficient run-length to obtain posterior inferences to the required level of precision. Because we pooled results of two independent chains, we based inferences on 100,000 draws.

The use of proposal densities that closely approximated the true posterior yielded similar rates of convergence even for models that required Metropolis-Hastings steps. For example, for $\text{SepStrat}_{\text{eig}}$ we used the posterior under the marginal uniform prior as a proposal density and obtained an acceptance rate of nearly 0.20.

5.2. Magnitude of Variances and Correlations

The posterior mean values of $E[\bar{\boldsymbol{\mu}} \mid \boldsymbol{Y}]$, $E[\bar{\boldsymbol{\sigma}} \mid \boldsymbol{Y}]$, and $E[\bar{\boldsymbol{\Omega}}_{\mathcal{A}} \mid \boldsymbol{Y}]$ ($\bar{x}$ is the average over the components of x) were converted to the $[0, 1]$ interval scale and compared between the approaches. Each approach obtains a very similar estimate of $E[\bar{\boldsymbol{\mu}} \mid \boldsymbol{Y}]$ and $E[\bar{\boldsymbol{\sigma}} \mid \boldsymbol{Y}]$. However, the expected mean correlation differed substantially across approaches, $E[\bar{\boldsymbol{\Omega}} \mid \boldsymbol{Y}] = 0.427$ (näive analysis described at the top of Section 5), 0.520 (ML), 0.555 (inverse-Wishart), 0.522 ($\text{SepStrat}_{\text{flat}}$), 0.520 ($\text{SepStrat}_{\text{vex}}$), 0.488 ($\text{SepStrat}_{\text{cave}}$), and 0.442 ($\text{SepStrat}_{\text{eig}}$). The differences among the approaches were consistent across response scales.

We computed the correlations between approaches of the components of $E[\boldsymbol{\Omega} \mid \boldsymbol{Y}]$. The lowest correlations (around 0.8) were between the näive approach and all other approaches. The next lowest correlations (around 0.96) were between the inverse-Wishart prior and the remaining approaches, while the highest correlations (above 0.99) were between $\text{SepStrat}_{\text{flat}}$, $\text{SepStrat}_{\text{vex}}$, and $\text{SepStrat}_{\text{cave}}$. These high correlations suggest that for this large sample from a large number of domains the likelihood function dominates the prior. When we re-ran the analysis with only half of the domains, and found that the correlations were smaller but similarly ordered.

In results not presented we found that as we increased the b_0 parameter of either the inverse-Wishart prior or the SepStrat priors, the magnitude of the correlations decreased. Both priors pull the covariance matrix toward a prior value with small off-diagonal elements, and so as the prior became more informative the posterior means of the correlations decreased in value.

5.3. Dimension of Factor Space

Following Guttman (1954) we defined the number of factors as the number of eigenvalues of the reduced correlation matrix Ω_A^{red} that exceed its mean eigenvalue (Table 1), obtaining a point estimate using maximum likelihood and a full posterior distribution with Bayesian approaches.

The number of factors selected ranged from 4 to 6 across all draws from the Bayesian models, with posterior means between 4 and 5 for each. The inverse-Wishart prior had a posterior mean of 4.6 and was one of two approaches ($\text{SepStrat}_{\text{eig}}$ was the other) that ever had 6 eigenvalues above the mean. The näive analysis estimated 6 factors, suggesting that the dimension of the factor space would be over-estimated if this approach was used for the analysis.

The posterior distributions of the number of factors under $\text{SepStrat}_{\text{flat}}$ and $\text{SepStrat}_{\text{vex}}$ were almost identical, with posterior means around 4.2. Under the $\text{SepStrat}_{\text{cave}}$ prior, the posterior mean was 4.5. Even though we deliberately favored 4 dimensions in specifying the prior, the mode of the posterior distribution for $\text{SepStrat}_{\text{eig}}$ was 5. Within the SepStrat families of priors the number of factors supported by the model increased with b_0.

We also fitted the eigenvalue prior model using the Dirichlet density with $\kappa_i = 0.9$ for $i \leq 6$ and $\kappa_i = 0.1$ for $i > 6$; that is, with more large eigenvalues and a larger prior ratio between large and small eigenvalues. The corresponding posterior probabilities of selecting five, six, and seven dimensions were higher than with the original Dirichlet prior.

5.4. Factor Structure

Table 2 presents certainty, instability, separation and discordance scores for each approach and pair of approaches when the number of factors is constrained to equal four. The certainty score ranged from 92.87 for the Inverse-Wishart model to just over 98 for $\text{SepStrat}_{\text{cave}}$ and $\text{SepStrat}_{\text{eig}}$, suggesting that these last two priors yield the least variable factor assignments. However, the instability score for $\text{SepStrat}_{\text{eig}}$ is higher than for the other models, implying that it had the most variable loadings. The instability scores for the other Bayesian models were fairly similar.

The highest separation score of 52.43 occurred under the $\text{SepStrat}_{\text{flat}}$ prior while the smallest separation score of 45.37 occurred with the näive analysis (just below that of $\text{SepStrat}_{\text{eig}}$). In general the separation score was lower for models (inverse-Wishart prior, $\text{SepStrat}_{\text{cave}}$, and $\text{SepStrat}_{\text{eig}}$) with higher posterior probabilities for 5 factors by the Guttman criterion.

The discordance scores quantify the distance between the factor loading matrices with higher scores implying greater distance (Table 2). The näive approach is least like the other approaches (scores around 98) and so likely supports a different factor structure. Among the Bayesian models $\text{SepStrat}_{\text{flat}}$ is the closest to the MLE (lowest discordance score) while the inverse-Wishart prior is the furthest from the MLE. Within the Bayesian models the lowest discordance score was obtained for $\text{SepStrat}_{\text{flat}}$ and $\text{SepStrat}_{\text{vex}}$ (8.59). As expected the discordance scores were larger when only half of the domains were used in the analysis. $\text{SepStrat}_{\text{eig}}$ was most similar to $\text{SepStrat}_{\text{cave}}$ since both tended to support large correlations and big eigenvalues.

5.5. CAHPS Composite Items: Relationship to Hypothesized Dimensions

Using the marginal uniform prior ($\text{SepStrat}_{\text{flat}}$) for $\mathbf{\Phi}$ we computed the posterior mean loading matrix assuming four factors with Promax rotation and Kaiser normalization. We assigned items to factors using the maximum loading rule for the 31 CAHPS report items (11 1–4, 11 1–3, and nine 1–2 items) that are usually summarized in composites.

To help interpret the remaining factors we ordered the items first by the dimensions of quality hypothesized when the survey was constructed (getting a doctor, doctor's office, doctor interactions, services and equipment, prescription drugs, customer service, and vacinations) and then within each dimension by the survey order. Two items, "interferes day-to-day" and "flu shots last year through plan" were not classified by these original groupings. Because the "flu shots last year through plan" item formed a distinct factor by itself, we excluded it as in previous analyses (Zaslavsky and Cleary 2002).

Items were assigned to factors using the maximum loading rule except that "happy with personal MD" was assigned to a factor with a slightly lower loading than its maximum loading to make the resulting factors easier to interpret. Table 3 shows the posterior means and standard deviations of the factor loadings and the posterior probabilities that each item has its largest loading on each factor.

The only item whose assignments to factors was ambiguous was "problem getting therapy", which had substantial loadings on the Doctor and Plan Customer Service factors and a non-trivial loading on the Vacinations factor. The "problem getting home health" and "interferes day-to-day" items each had a high loading on one factor and a moderately high loading on a second factor. The decisiveness of the reported factor structure was supported by the relatively small posterior standard deviations of the factor loadings.

With the exception of services and equipment, and prescription drugs, the originally hypothesized dimensions remained intact. However, only three distinct factors were apparent. The three groups of doctor ratings formed a general doctor factor with part of the services and equipment and prescription drugs groups of items. Customer service combined with the remaining services and equipment and prescription drugs items to form a Plan Customer Service factor. The vacination items grouped together and dominate the Vacinations factor.

The fact that fewer factors were obtained than hypothesized is not surprising since CAHPS surveys are often designed using respondent-level analyses. Thus the hypothesized domains might be more reflective of relationships at the respondent level than the domain level. There might be a single dimension of quality at the domain level related to doctors, with finer groupings of the items obtained from respondent-level analyses attributable to individual variation. Similarly, in the CAHPS hospital study pilot survey, where nine dimensions of care were hypothesized, seven similar factors were found at the patient level, but only three were supported at the hospital level (O'Malley et al. 2005).

The factor structure in Table 3 has high face validity. The services and equipment items involve both the doctor and the health plan so it is not surprising that they load on both Doctor and Plan Customer Service. The services and equipment item most related to the health plan, "plan provided all help needed", loaded on Plan Customer Service. The inclusion of the "long wait past appointment" item with Vacinations suggests that flu and pneumonia shots are best encouraged when doctors' offices are well-organized to cope with the flow of patients.

It was surprising that the prescription drugs items did not hang together, although they only appeared together in the survey for three of the five years. However, there is some logic to the way in which they grouped; “problem getting Rx from plan”, which is purely a function of the plan, was aligned with Plan Customer Service, whereas “get prescription through plan”, which involves the doctor (to get the prescription) was understandably aligned with Doctor.

Because the certainty score is perfectly correlated with the maximum posterior probability for an item-factor assignment, it is not presented in Table 3. The large number of posterior probabilities of factor assignments close to 1 or 0 give us confidence in the factor structure.

5.6. CAHPS Composite Items: Relationship to Rating Items

The factor analysis at the plan level guides us in grouping items for reports. The sum of the scores for the items assigned to a factor, or composite score, is reported to summarize the individual scores. For comparability among items with different scales, items were rescaled to the $[0, 1]$ interval when computing composites. The items in each factor are indicated by the boldface factor loadings in Table 3.

Because the four rating items (personal doctors, specialist, care, plan) are considered to be summary judgements influenced by experiences described more specifically in the report items, regressing them on the scores for the three composites sheds light on the face validity of the composite items and helps in their interpretation. We first rescaled the ratings and composites to have variance 1 so the resulting regression coefficients β_{comp} are standardized regression coefficients, where the regression coefficient on the jth composite for the ith rating is $\beta_{\text{comp},ij}$. .

For each draw of Σ, the least squares parameters are computed from $\boldsymbol{\Sigma}$ and the matrix $\boldsymbol{U}$ whose ith row contains the item weights for the i'th composite. Let $\boldsymbol{\Sigma}_{11}$, $\boldsymbol{\Sigma}_{22}$ and $\boldsymbol{\Sigma}_{12}$ denote the blocks of $\boldsymbol{\Sigma}$ corresponding to the rating items, the report items, and the covariances of the rating and report items. The variances of the composites and the covariance between the ratings and the composites are given by $\boldsymbol{C}_{22} = \boldsymbol{U}\boldsymbol{\Sigma}_{22}\boldsymbol{U}^T$ and $\boldsymbol{C}_{12} = \boldsymbol{\Sigma}_{12}\boldsymbol{U}^T$ respectively, and then $\boldsymbol{\beta}_{\text{comp}} = \boldsymbol{C}_{12}\boldsymbol{C}_{22}^{-1} = \boldsymbol{\Sigma}_{12}\boldsymbol{U}^T(\boldsymbol{U}\boldsymbol{\Sigma}_{22}\boldsymbol{U}^T)^{-1}$. The significance of the coefficients is assessed by $P_{ij} = \min\{\text{pr}(\beta_{\text{comp},ij} > 0 \mid \boldsymbol{Y}), \text{pr}(\beta_{\text{comp},ij} < 0 \mid \boldsymbol{Y})\}$, which functions like a p-value for a two-sided hypothesis test in the sense that smaller values indicate higher significance.

Table 4 reports the posterior means of β_{comp}. The doctor composite, which includes all the doctor related items is the strongest predictor of the personal doctors, specialist, and care rating items. The plan customer service composite is an extremely strong predictor of the plan rating but has much lower associations with the other rating items. The vacinations composite was moderately predictive of the personal doctor rating, moderately predictive of the care rating, and only weakly predictive of the specialist or plan ratings; vacinations are not typically provided by specialists or plans, but they are a component of the care provided to the respondent and are often provided in the nonspecialist doctor's office.

6. Conclusions

We have presented statistical methodology for estimating a domain-level covariance matrix of item means in the presence of nonresponse under a hierarchical model, using a generalized covariance function to assist in estimation of respondent-level covariance parameters. We compared maximum likelihood estimates to Bayesian inferences under several different priors.

This methodology has many applications in healthcare assessment surveys. For example, a similar method was used to analyze pilot data from the CAHPS hospital survey (O'Malley et al. 2005). The results were essential to developing composite reports of the quality of care and services provided by hospitals, and to streamlining the survey by identifying items that could be omitted without losing the ability to measure each identified dimension of quality. These analyses were conducted by maximum likelihood and were simplified by ignoring the dependence of $V(\boldsymbol{\theta}_h)$ on $\boldsymbol{\theta}_h$ in the E-step, enabling derivation of the full conditional posterior distribution of $\boldsymbol{\theta}_h$ and use of the standard EM algorithm. The Bayesian methodology developed here will be used for future CAHPS applications. Using appropriate estimates of sampling covariance matrices, the same method can be applied to calculating between-unit covariances of subgroup means, such as mean responses calculated separately for sick and healthy members of each health plan (Zaslavsky and Cleary 2002), or domain-specific regression coefficients (Zaslavsky, Zaborski, and Cleary 2000b). We anticipate that a similar approach could be used to analyze survey data that assesses other types of institutions such as schools. The methodology can also be extended to allow covariates to be incorporated at both the individual and unit levels.

For the Bayesian models we focused on the effect of the prior for $\boldsymbol{\Sigma}$. Because improper default priors may yield improper posteriors, we only considered proper priors. As expected, in models with higher prior precision the posterior distributions of correlations were pulled more toward the prior, which typically favored correlations closer to zero than suggested by the data. The Dirichlet prior on the eigenvalues of the correlation matrix yielded a posterior that supported more factors than with the other priors and as a consequence the mean loadings were lower.

The Bayesian models considered in this paper were all derived from the same matrix decomposition of $\boldsymbol{\Sigma}$. Alternative matrix decompositions, including the Cholesky decomposition of $\boldsymbol{\Sigma}$ (e.g., Pinheiro and Bates 1996) or its inverse (e.g., Pourahmadi 1999, Liu 1993) and the spectral decomposition of $\mathbf{\Sigma}$ including the matrix logarithm (e.g., Leonard and Hsu 1992; Chiu, Leonard, and Tsui 1996), often involve parameterizations that less intuitively express prior information about correlations. The "separation strategy" priors are also more convenient for Bayesian computation.

Because the Bayesian models allowed a wide range of inferences to be evaluated, particularly concerning the variability of the results of the exploratory factor analysis, they are suitable for drawing conclusions about the underlying factor structure. Examples of inferences facilitated by the Bayesian models that are not easily derived using MLE include evaluating the variability in descriptive parameters (i.e., those based on some decision rule) for the number of underlying factors, and deriving the probability that an item is assigned to a given factor.

We use the fitted model in a descriptive way, and therefore have not emphasized model checking. Bayesian predictive p-values could be used to determine how well the fitted model reproduces the observed data and thus the appropriateness of the model (Gelman, Carlin, Stern, and Rubin 1995, pp. 169-174). Because the focus of our analysis is Σ one might consider constructing Bayesian predictive p-values for statistics that reflect features of Σ through the sample correlation matrix.

In future work we plan to develop fully Bayesian methodology to test hypotheses about the structure of a domain-level covariance matrix by incorporating the factor structure of the items into the prior. This will embed the entire analysis within a single model allowing both confirmatory and exploratory hypotheses to be formally tested.

References

Barnard, J., McCulloch, R., and Meng, X.-L. (2000), "Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage," *Statistica Sinica*, 10, 1281–1311.

Best, N., Cowles, M. K., and Vines, K. (1995), *Convergence diagnosis and output analysis software for Gibbs sampling output.*, MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 2SR, UK.

Brooks, S. P. and Gelman, A. (1998), "General methods for monitoring convergence of iterative simulations," *Journal of Computational and Graphical Statistics*, 7, 434–455.

Chiu, T. Y. M., Leonard, T., and Tsui, K.-W. (1996), "The matrix-logarithmic covariance model," *Journal of the American Statistical Association*, 91, 198–210.

Crofton, C., Lubalin, J. S., and Darby, C. (1999), "The consumer assessment of health plans study (CAHPS): Foreword," *Medical Care*, 37, MS1–MS9.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B, Methodological*, 39, 1–22.

Fay, R. E. and Herriot, R. A. (1979), "Estimates of income for small places: An application of James-Stein procedures to census data," *Journal of the American Statistical Association*, 74, 269–277.

Gelman, A. (2005), "Prior distributions for variance parameters in hierarchical models," *Bayesian Data Analysis*, 1, 1–19.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995), *Bayesian data analysis*, London: Chapman & Hall Ltd.

Gelman, A., Roberts, G. O., and Gilks, W. R. (1996), "Efficient Metropolis jumping rules," in *Bayesian Statistics 5 – Proceedings of the Fifth Valencia International Meeting*, 599–607.

Ghosh, M. and Rao, J. N. K. (1994), "Small area estimation: An appraisal," *Statistical Science*, 9, 55–76.

Goggins, W. B., Finkelstein, D. M., Schoenfeld, D. A., and Zaslavsky, A. M. (1998), "A Markov chain Monte Carlo EM algorithm for analyzing interval-censored data under the Cox proportional hazards model," *Biometrics*, 54, 1498–1507.

Goldstein, E., Cleary, P. D., Langwell, K. M., Zaslavsky, A. M., and Heller, A. (2001), "Medicare Managed Care CAHPS: A tool for performance improvement," *Health Care Financing Review*, 22, 101–107.

Guttman, L. (1954), "Some necessary conditions for common-factor analyses." *Psychometrika*, 19, 149–161.

Hastings, W. K. (1970), "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, 57, 97–109.

Heidelberger, P. and Welch, P. D. (1983), "Simulation run length control in the presence of an initial transient," *Operations Research*, 31, 1109–1144.

Hellmich, M., Abrams, K. R., Jones, D. R., and Lambert, P. C. (1998), "A Bayesian approach to a general regression model for ROC curves." *Academic Radiology*, 18, 436–443.

Leonard, T. and Hsu, J. S. J. (1992), "Bayesian inference for a covariance matrix," *The Annals of Statistics*, 20, 1669–1696.

Liu, C. (1993), "Bartlett's decomposition of the posterior distribution of the covariance for normal monotone ignorable missing data," *Journal of Multivariate Analysis*, 46, 198–206.

McLachlan, G. J. and Krishnan, T. (1997), *The EM algorithm and extensions*, New York: John Wiley & Sons.

Meng, X.-L. and Zaslavsky, A. M. (2002), "Single observation unbiased priors," *Annals of Statistics*, 30, 1345–1375.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953), "Equations of state calculations by fast computing Machines," *Journal of Chemical Physics*, 21, 1078–92.

O'Malley, A. J. and Zaslavsky, A. M. (2005), "Variance-covariance functions for domain means of ordinal survey items," *Survey Methodology*, 31, 169–182.

O'Malley, A. J., Zaslavsky, A. M., Hays, R. D., Hepner, K., Keller, S., and Cleary, P. D. (2005), "Exploratory factor analyses of the CAHPS® hospital pilot survey responses across and within medical, surgical, and obstetric services," *Health Services Research*, 40, 2078–2095.

Pinheiro, J. C. and Bates, D. M. (1996), "Unconstrained parameterizations for variance-covariance matrices," *Statistics and Computing*, 6, 289–296.

Pourahmadi, M. (1999), "Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation," *Biometrika*, 86, 677–690.

Schnaier, J. A., Sweeny, S. F., Williams, V. S., Kosiak, B., Lubalin, J. S., Hays, R. D., and Harris-Kojetin, L. D. (1999), "Special issues addressed in the CAHPS survey of Medicare managed care beneficiaries," *Medical Care*, 37, MS69–MS78.

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

Zaslavsky, A. M., Beaulieu, N. D., Landon, B. E., and Cleary, P. D. (2000a), "Dimensions of consumer-assessed quality of Medicare managed-care health plans," *Medical Care*, 38, 162–174.

Zaslavsky, A. M. and Cleary, P. D. (2002), "Dimensions of plan performance for sick and healthy members on the Consumer Assessments of Health Plans Study 2.0 survey," *Medical Care*, 40, 951–964.

Zaslavsky, A. M., Zaborski, L., and Cleary, P. D. (2000b), "Does the Effect of Respondent Characteristics on Consumer Assessments Vary Across Plans?" *Medical Care Research and*

Review, 57, 379–394.

Table 1: The number of eigenvalues greater than the mean eigenvalue of the reduced correlation matrix for the domain-level correlation matrix.

	Approach						
	Näive	Maximum	Inverse	Separation Strategy			
	Analysis	Likelihood	Wishart	Flat	Convex	Concave	Eigenvalue
Estimate/Posterior Mean	6	4	4.61	4.18	4.22	4.51	4.97
Pr(4 factors)	n/a	n/a	0.41	0.82	0.78	0.49	0.04
Pr(5 factors)	n/a	n/a	0.57	0.18	0.22	0.51	0.95
Pr(6 factors)	n/a	n/a	0.02	0.00	0.00	0.00	0.01

Table 2: Certainty, Instability, Separation and Discordance scores for the domain-level unrotated factor-loading matrix for Näive analysis, MLE, and the Bayesian models, assuming 4 factors.

	Approach						
	Näive	Maximum	Inverse	Separation Strategy			
Quantity/Approach	Analysis	Likelihood	Wishart	Flat	Convex	Concave	Eigenvalue
Certainty Score							
Estimate	n/a	n/a	92.87	96.74	96.90	98.16	98.10
Instability Score							
Within Method (W)	n/a	n/a	0.139	0.120	0.137	0.130	0.179
Separation Score							
Estimate/Mean	45.37	51.65	47.84	52.43	52.17	50.70	46.76
Standard Deviation	n/a	n/a	1.935	1.605	1.739	1.729	1.302
Discordance Score							
Näive Analysis			98.01	98.06	97.78	97.87	97.10
Maximum Likelihood			75.79	54.68	57.10	67.56	67.07
Inverse Wishart				79.77	80.02	83.55	83.63
Separation Strategy: Flat					8.59	27.61	48.92
Separation Strategy: Convex						19.68	43.03
Separation Strategy: Concave							21.58

The Certainty, Instability and Separation scores are computed for each approach, while the Discordance score is computed for each pair of approaches that involve at least one Bayesian approach.

Table 3: Analysis of factor loadings for $SepStrat_{flat}$ using oblique promax rotation.

			Doctor			Plan Customer Service			Vacinations		
Grouping	Name	Type	Mn	SD	Pr	Mn	SD	Pr	Mn	SD	Pr
Getting a Doctor	Happy with personal MD	1–3	**0.39**	0.04	0	0.59	0.04	1	-0.17	0.04	0
Getting a Doctor	Problem getting referral	1–3	**0.60**	0.03	1	0.18	0.04	0	0.25	0.04	0
Getting a Doctor	Doctor knows important facts	1–2	**0.68**	0.04	1	0.19	0.06	0	-0.14	0.06	0
Getting a Doctor	MD understand affect	1–2	**0.59**	0.04	1	0.03	0.06	0	-0.04	0.06	0
Doctor's Office	Get advice from MD office	1–4	**0.89**	0.02	1	0.00	0.02	0	0.09	0.03	0
Doctor's Office	Routine care soon as wanted	1–4	**1.02**	0.01	1	-0.16	0.03	0	-0.21	0.03	0
Doctor's Office	Care for illness quickly	1–4	**0.81**	0.02	1	0.04	0.03	0	0.04	0.04	0
Doctor's Office	Long wait past appointment	1–4	0.19	0.05	0	0.00	0.05	0	**0.63**	0.05	1
Doctor's Office	MD courtesy and respect	1–4	**0.73**	0.03	1	-0.04	0.03	0	0.32	0.04	0
Doctor's Office	MD helpful	1–4	**0.85**	0.02	1	0.02	0.02	0	0.16	0.03	0
Doctor's Office	Get needed care	1–3	**0.76**	0.03	1	0.01	0.03	0	0.25	0.04	0
Doctor's Office	Delays in getting care	1–3	**0.47**	0.03	0.945	0.38	0.03	0.055	0.24	0.03	0
Doctor Interactions	MD listens carefully	1–4	**0.96**	0.01	1	0.02	0.02	0	-0.07	0.03	0
Doctor Interactions	MD explains things	1–4	**0.90**	0.02	1	0.02	0.03	0	-0.01	0.03	0
Doctor Interactions	MD shows respect	1–4	**0.94**	0.01	1	0.02	0.03	0	-0.04	0.03	0
Doctor Interactions	MD spends enough time	1–4	**0.99**	0.01	1	-0.05	0.02	0	-0.08	0.03	0
Services & Equipment	Get special med equipment	1–3	0.29	0.04	0	**0.44**	0.04	0.995	0.26	0.04	0.005
Services & Equipment	Problem get therapy	1–3	**0.36**	0.04	0.495	0.35	0.04	0.365	0.32	0.04	0.14
Services & Equipment	Problem get home hlth	1–3	**0.35**	0.04	0.855	0.20	0.05	0.015	0.28	0.05	0.13
Services & Equipment	Plan provide all help needed	1–2	0.22	0.04	0	**0.51**	0.04	1	0.09	0.05	0
Prescription Drugs	Problem get Rx from plan	1–3	-0.17	0.04	0	**0.68**	0.04	1	0.05	0.05	0
Prescription Drugs	Get prescription through plan	1–2	**0.49**	0.05	0.985	-0.17	0.05	0	0.29	0.05	0.015
Customer Service	Customer service helpful	1–4	0.24	0.03	0	**0.79**	0.03	1	-0.04	0.03	0
Customer Service	Problem get info	1–3	-0.19	0.03	0	**0.89**	0.03	1	-0.03	0.03	0
Customer Service	Problem get help on call	1–3	0.21	0.03	0	**0.82**	0.03	1	-0.06	0.03	0
Customer Service	Problem with paperwork	1–3	-0.03	0.03	0	**0.86**	0.03	1	0.01	0.03	0
Customer Service	Complaint or prob with plan	1–2	-0.15	0.03	0	**0.80**	0.03	1	0.16	0.03	0
Vacinations	Flu shot last year	1–2	-0.04	0.03	0	0.16	0.04	0	**0.79**	0.03	1
Vacinations	Ever have a Pneumonia shot	1–2	-0.17	0.03	0	0.05	0.04	0	**0.89**	0.03	1
Not Classified	Interferes day-to-day	1–2	**0.34**	0.04	0.79	-0.06	0.05	0	0.27	0.06	0.21
Excluded	Flu shot last yr through plan	1–2									

Note: Items are ordered by their hypothesized composites.

Table 4: Regression of CAHPS-MMC rating items on composite items.

	Rating Item							
	Personal Doctors		Specialist		Care		Plan	
Composite	$\boldsymbol{\beta}_{\text{comp},ij}$	P_{ij}	$\boldsymbol{\beta}_{\text{comp},ij}$	P_{ij}	$\boldsymbol{\beta}_{\text{comp},ij}$	P_{ij}	$\boldsymbol{\beta}_{\text{comp},ij}$	P_{ij}
Doctor	1.12	0.00	1.15	0.00	1.18	0.00	0.096	0.00
Plan Customer Service	0.015	0.36	-0.133	0.00	0.107	0.00	1.11	0.00
Vacinations	-0.362	0.00	-0.055	0.06	-0.176	0.00	-0.085	0.00

$\boldsymbol{\beta}_{\text{comp},ij}$ is the posterior mean of the standardized regression coefficients for the regression of the rating items on the composite items, while $P_{ij} = \min\{\text{pr}(\beta_{\text{comp},ij} > 0 \mid \boldsymbol{Y}), \text{pr}(\beta_{\text{comp},ij} < 0 \mid \boldsymbol{Y})\}$.

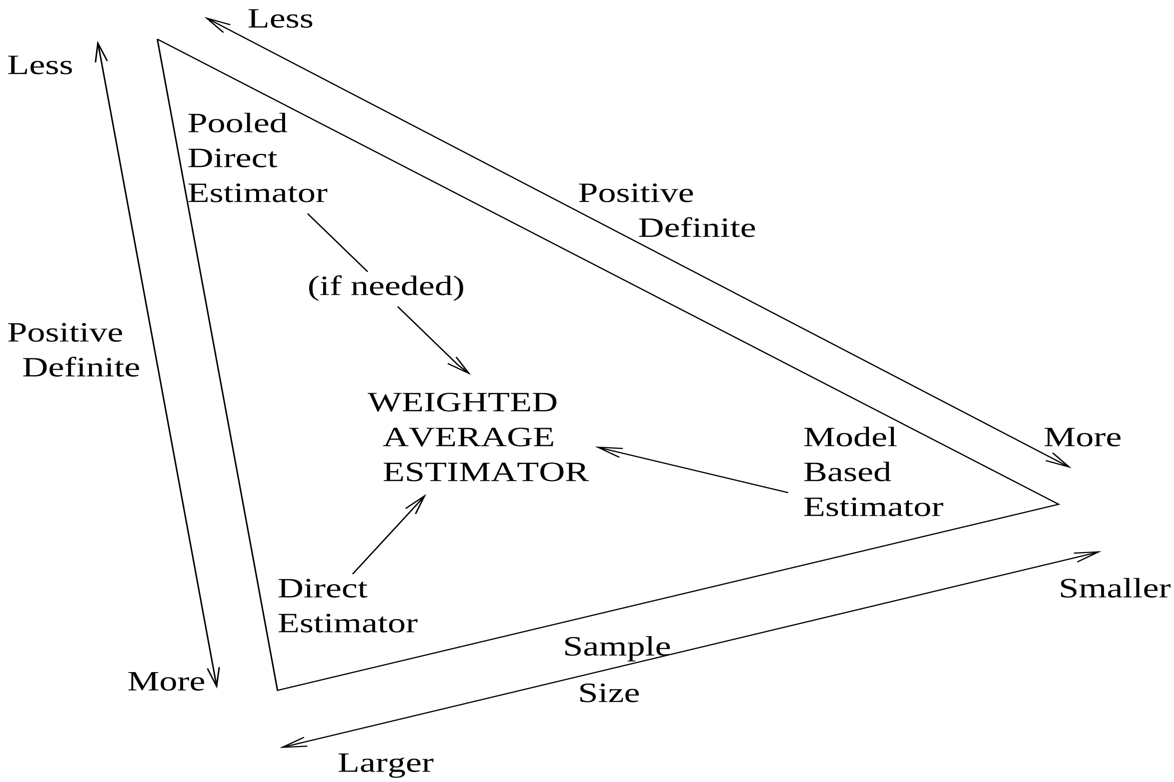

Figure 1: Diagram illustrating how the direct and model-based covariance within domain covariance estimators are combined to form a positive definite covariance matrix.