

# Nonparametric Bayesian Multiple Imputation for Missing Data Due to Mid-study Switching of Measurement Methods

Lane F. Burgette and Jerome P. Reiter\*

January 22, 2011

## Abstract

Investigators often change how variables are measured during the middle of data collection, for example in hopes of obtaining greater accuracy or reducing costs. The resulting data comprise sets of observations measured on two (or more) different scales, which complicates interpretation and can create bias in analyses that rely directly on the differentially measured variables. We develop multiple approaches for handling mid-study changes in measurement for settings in the absence of calibration data, i.e., no subjects are measured on both (all) scales. This setting creates a seemingly insurmountable problem for multiple imputation: since the measurements never appear jointly, there is

---

\*Lane F. Burgette is Postdoctoral Research Associate ([1b131@stat.duke.edu](mailto:1b131@stat.duke.edu)) and Jerome P. Reiter ([jerry@stat.duke.edu](mailto:jerry@stat.duke.edu)) is Mrs. Alexander Hehmeyer Associate Professor, Department of Statistical Science, Duke University, Durham, NC 27708-0251. The authors wish to thank Howard Chang, Sharon Edwards, Marie Lynn Miranda, and Geeta Swamy for helpful comments. This research was funded by Environmental Protection Agency grant R833293.

no information in the data about their association. We resolve the problem by making an often scientifically reasonable assumption that each measurement regime accurately ranks the samples but on differing scales, so that, for example, an individual at the  $q$ th percentile on one scale should be at about the  $q$ th percentile for the other scale. We use rank-preservation assumptions to develop three imputation strategies that flexibly transform measurements made in one measurement scale to measurements made in another: an MCMC-free approach based on permuting ranks of measurements, and two approaches based on dependent Dirichlet process mixture models for imputing ranks conditional on covariates. We use simulations to illustrate conditions under which each strategy performs well, and present guidance on when to apply each. We apply these methods to a large study of birth outcomes in which investigators collected mothers' blood samples to measure levels of environmental contaminants. Mid-way through data ascertainment, the study switched from one analytical laboratory to another, and it is clear that the two labs report the measurements according to different measurement scales. We use nonparametric Bayesian imputation models to obtain sets of plausible measurements on a common scale, and estimate quantile regressions of birth weight on various environmental contaminants. The results suggest that blood lead levels interact with smoking status in ways not previously reported in the epidemiologic literature.

Keywords: Dirichlet process, Gaussian process, Permutation, Rank

## 1 INTRODUCTION

In large-scale data collections, it is not uncommon for the investigators to switch measurement procedures during the data collection phase. As examples, investigators

collecting biomedical data may switch assay labs or instruments to reduce costs or improve accuracy; and, investigators running prospective studies may change question wording or survey mode for some variables. Hence, at the end of collection, the data comprise some participants measured one way and others a different way. When the two (or more) measurement scales differ, inferences based on the combined data can be inaccurate and difficult to interpret.

It is relatively straightforward to adjust for differing scales when investigators can measure subsets of data subjects on the multiple scales. For example, one can use missing data methods to create plausible values of all measurements (Schenker and Parker, 2003; Cole et al., 2006; Durrant and Skinner, 2006; Thomas et al., 2006), and analyze the imputed data using the preferred measurement scales. Sometimes, however, it is not practical or feasible to measure data subjects on more than one scale simultaneously. When faced with this situation for numerical measurements, analysts often use the simple approach of standardizing the measurements to get them on a common scale. However, a one unit change on some scale may mean something different on another scale, and the extent of that difference may change for low and high levels of the measured variable. Furthermore, standardizing fails when background characteristics related to the measured variable differ across measurement groups. Another approach is to delete all but the preferred measurements, and use missing data methods on the remaining data. This sacrifices potentially useful information in the measurements, leading to inefficient inferences.

In this article, we present three alternative strategies for handling mid-study changes in measurement for numerical data. We present the strategies for two measurement scales; the methods easily extend to more than two scales. To aid description, we define the destination scale to comprise the values after the mid-study change in measurement, and define the source scale to comprise the initial measurements.

The key assumption underlying the approaches is that rankings are roughly preserved across the measurement scales; e.g., if an individual is at the 10th percentile on the source scale, she should be at about the 10th percentile of the destination scale. Such assumptions are reasonable in many settings. For example, the procedures used by two assay labs may report different levels of some agent, but it may be biologically sensible to assume that someone who measures high (low) by one procedure would measure high (low) by the other procedure. Using only rank-preservation assumptions, it is possible to impute the missing destination scale measurements for source-scale records, either as part of parameter estimation in Bayesian models or as part of a multiple imputation analysis. We pursue the latter here.

The three methods can be ordered based on the extent to which they make use of information about covariates related to the differentially measured variables for imputation. The first method, which we call the rank permutation (RP) method, involves imputing the destination ranks—and subsequently the destination values—of the measurements in the source data independently of covariates. The second method, which we call rank-preserving prediction (RPP), involves imputing the destination values of the measurements in the source data while taking covariate information into account and maintaining the observed within-lab rankings. The third method, which we call matched conditional quantiles (MCQ), equates conditional quantiles in density regressions of the values in each measurement scale. Roughly, if an observation is at the  $q$ th conditional quantile in one scale, MCQ imputes it at the  $q$ th conditional quantile in the other scale. MCQ ensures that ranks from the source data are preserved locally with respect to the space of the covariates, whereas RPP ensures that ranks from the source data are preserved globally.

For both RPP and MCQ, we estimate conditional densities using nonparametric Bayesian approaches based on dependent Dirichlet process mixture models (MacEach-

ern, 1999; De Iorio et al., 2004; Fronczyk and Kottas, 2010). These flexible models are advantageous for imputation, since they enable the analyst to relate two measurement scales using minimal assumptions while controlling for relevant background characteristics. RP involves simple permutations of observed ranks and so is less computationally demanding, which may make it more appealing to analysts than RPP and MCQ in some settings.

The remainder of the paper is arranged as follows. In Section 2, we describe the motivating example for the development of these methods, namely a mid-study change in assay labs during a prospective study of the relationships between environmental exposures and adverse birth outcomes. In Section 3, we describe the three proposed methods in the context of two measurement scales. In Section 4, we present results of simulation studies that illustrate the methods and illuminate conditions under which each performs well. In Section 5, we apply the results on the motivating example, with a focus on mothers' blood lead concentrations that were made by different assay labs. Finally, in Section 6, we conclude with a brief discussion of broad applications of these methodologies.

## **2 MOTIVATING EXAMPLE: THE HEALTHY PREGNANCY, HEALTHY BABY STUDY**

The Healthy Pregnancy, Healthy Baby Study (HPHBS) is an ongoing observational cohort study that is focused on the etiology of adverse birth outcomes. The intent of the study is to investigate how environmental, social, and host factors are related to outcomes like birth weight and gestational age at birth. Since July 2005, the study has recruited women aged 18 and up who are pregnant with a singleton gestation. These expectant mothers are recruited at the Duke University Obstetrics Clinic and the

Durham County Health Department Prenatal Clinic, both of which are in Durham, NC (<http://epa.gov/ncer/childrenscenters/duke.html>). As of this analysis, the data comprise 1435 non-Hispanic black and white women who have given birth.

The study investigators collect blood samples from the expectant mothers to measure their exposures to the pollutants lead, mercury, cadmium, and cotinine. In the third year of data collection, the investigators switched from one analytical lab to another that promised finer assay resolution. However, after enough samples were taken from the new lab, the investigators noticed that the marginal distributions of the pollutants' concentrations differed greatly between the two labs in ways not explainable solely by the differing degrees of coarseness of the reported values. For example, Figure 1 displays the two sets of marginal concentrations for mothers' blood lead levels, broken down by race. If one scale were a rounded version of the other, we would expect more of the mothers to be above the detection limit in the coarser scale; or, we would expect the finer scale's values to rise less quickly. On the other hand, the coarser scale has more high values than we would expect if it were a rounded version of the finer scale.

The differences in marginal distributions could result from differences in the background characteristics of the samples between labs. In fact, a greater percentage of mothers identify their race as white in the lab with finer assay resolution compared to those who were measured in the original lab. However, within racial groups, other characteristics of study participants are not appreciably different for the cohorts measured in the two labs: logistic regressions of lab assignment on a function of tobacco use, age, and birth weight yield no coefficients that are significant at the 10% level. Therefore, we attribute within-race differences in the marginal distributions to differences in the two labs' measurement methods.

Due to the difficulties of getting blood samples from pregnant women and the

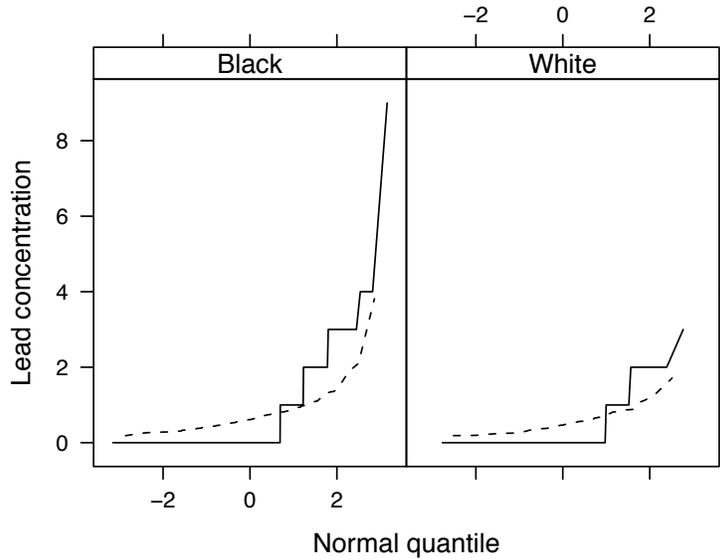


Figure 1: Normal quantile/quantile plots of lead data from the two labs, by mother's race.

need to preserve as much sample as possible for various assays, no mothers were measured in both labs. However, it is reasonable to assume that the two labs have low measurement errors but differing intrinsic scales; that is, each lab can properly rank samples (perhaps up to ties). Put another way, we assume that if the true value of an assay were  $y$ , lab 1 would report  $f_1(y)$  and lab 2 would report  $f_2(y)$  where  $f_1$  and  $f_2$  are increasing—but perhaps quite complicated—unknown functions. We do not have any information to determine whether  $f_1$  or  $f_2$  is the identity function. Hence, the best that we can do with the HPHBS data is to create a coherent scale for the measurements across samples. We cannot claim that to create imputed values that are in some true scale.

The HPHBS is a large study with many investigators analyzing the data. Hence, we adopt a multiple imputation approach (Rubin, 1987) to impute plausible values of

the pollutants on a coherent scale defined by the finer-resolution measurements. In particular, we use the methods described in Section 3 to create ten completed datasets so that each mother has either an actual concentration measurement (if she was measured by the finer-resolution lab) or a set of ten imputed concentration values (if she was measured by the original, coarser-resolution lab or not measured at all). With these completed datasets, investigators can use complete-data techniques on each imputed dataset, and combine results using simple rules (Reiter and Raghunathan, 2007).

### 3 DESCRIPTION OF THE METHODS

We now describe the rank permutation (RP), rank-preserving prediction (RPP), and matched conditional quantiles (MCQ) methods. Let  $Y$  represent the variable measured on two different scales, and let  $X$  represent all other variables in the dataset. We suppose that the values of  $Y$  observed in the source scale,  $y_{is}$  where  $i = 1, \dots, n_s$ , are ordered from smallest to largest, as are the values of  $Y$  observed in the destination scale,  $y_{id}$  where  $i = 1, \dots, n_d$ . Let  $\mathbf{y}_s$  and  $\mathbf{y}_d$  be the vectors of all individuals' observed data in the source and destination scales, respectively. Let  $\mathbf{y}_c$  denote the complete set of  $n_c = n_s + n_d$  observations in the destination scale. Note that elements of  $\mathbf{y}_c$  are observed for records in the destination-scale data but missing for records in the source-scale data.

#### 3.1 Rank Permutation

We begin with RP, which does not explicitly include covariate information in the imputation process and is simplest to implement computationally. RP relies on the factorization  $p(\mathbf{y}_c | \mathbf{y}_s, \mathbf{y}_d) = p(\mathbf{y}_c | \mathbf{r}_c, \mathbf{y}_s, \mathbf{y}_d) p(\mathbf{r}_c | \mathbf{y}_s, \mathbf{y}_d)$ , where  $\mathbf{r}_c$  is the unobserved set

of ranks of  $\mathbf{y}_c$ . If the elements of  $\mathbf{y}_c$  are assumed to be drawn independently from some common distribution, then  $p(\mathbf{r}_c|\mathbf{y}_s, \mathbf{y}_d)$  can be sampled as follows. Imagine an urn with  $n_s$  red balls for the source-scale observations and  $n_d$  blue balls for the destination-scale observations. Sample all  $n_c$  balls without replacement, numbering each ball after it is drawn with consecutive numbers from 1 to  $n_c$ . The numbers on the red balls are a draw of the ranks of the source-scale measurements if they were transformed into the destination scale.

For example, suppose that  $n_s = 3$  and  $n_d = 2$ . A drawn sequence from the urn might be  $B_1R_2R_3B_4R_5$ , with  $B$  for blue and  $R$  for red. The observed destination values are retained, so that  $y_{1c} = y_{1d}$  and  $y_{4c} = y_{2d}$ . We would sample—according to some distributional estimator applied to  $\mathbf{y}_d$ —imputed values of  $y_{2c}$  and  $y_{3c}$  so that  $y_{1c} < y_{2c} < y_{3c} < y_{4c}$ . Similarly, we sample  $y_{5c}$  restricted to be larger than  $y_{4c}$ . For simplicity, we draw from a discretized version of a Gaussian kernel density estimate, as implemented in the `density()` function in R (Venables and Ripley, 2002; R Development Core Team, 2010). This is a Monte Carlo technique, but not MCMC, so there are no significant computational concerns. An R implementation is available from the authors.

To illustrate the RP method, we consider the following data-generating setup. The marginal distribution of the  $n_d = 500$  destination measurements is standard normal. We transform from destination to source measurements using  $f(y) = -2.5 + 5 \exp\{-.5 + .2y\}$ . We then apply RP to impute plausible values of the  $n_s = 200$  source scale measurements in the destination scale. Figure 2 shows the marginal distribution of  $\mathbf{y}_c$  for ten realizations of RP. The imputed distributions are centered around  $\mathbf{y}_d$  with uncertainty comparable to the difference between  $\mathbf{y}_d$  and the source-scale observations after transformation by the true inverse of  $f$ . Because RP only uses the ranks of the source lab observations, we note that any strictly increasing

function  $f$  would yield qualitatively similar results.

It is possible to incorporate some auxiliary information by stratifying the observations according to covariates, and performing RP within each stratum. This approach can produce imputed values that do not respect the within-lab marginal ranks. It also can increase variance when sample sizes are small in some strata.

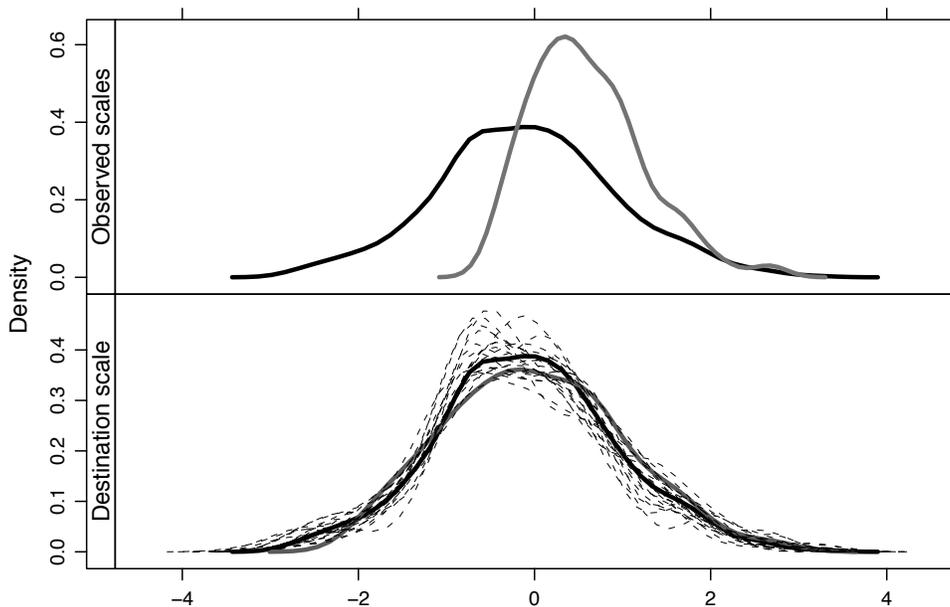


Figure 2: Example of the rank permutation (RP) method. The top panel displays density histograms of the observations from the source lab (gray) and destination lab (black). In the lower panel, the true density histogram of the transformed values is the gray line. Ten realizations of the RP method are displayed (thin dashed lines), along with the observed destination lab measurements (solid black).

### 3.2 Rank-Preserving Predictions

RPP is a natural extension to RP, as it gives priority to preserving the observed rankings for the source-scale records. The key modification is that RPP overcomes

the lack of covariate information in the RP approach, which for many settings would be problematic. For example, average blood lead levels tend to be higher for older women. This is partially due to a cohort effect: environmental lead exposure in the U.S. is lower now than it was several decades ago, with reductions in lead-containing paint and the 1996 ban on leaded gasoline (Thomas, 1995; Jacobs and Nevin, 2006). There is also an age effect, because lead accumulates in the skeletal system over the life course, with some of the stored lead being released during pregnancy (Gulson et al., 1999). If the women measured on the destination scale are mostly older than the women measured on the source scale, using RP could impute younger women with high ranks in the source scale to have lead values comparable to those for older women in the destination scale, which would not be appropriate.

To implement RPP, we estimate the conditional distribution of  $\mathbf{y}_c$  given covariates  $\mathbf{x}_c$  using the destination-scale data. For each source-scale record  $i$ , we sample a value of  $y_{ic}$  from this conditional distribution with the constraint that the rank of  $y_{ic}$  among all source records' ranks must be preserved; for example, if  $y_{is}$  was at the 20th percentile among source records, then its imputed  $y_{ic}$  should be at the 20th percentile among the imputed values for all source records.

More formally, the imputation proceeds as follows. We estimate the conditional distributions via a Bayesian density regression fit with the observed destination data, as described in Appendix A. In particular, we use a dependent Dirichlet process (DDP) model to capture the distribution of  $\mathbf{y}_d$  across the observed covariate space  $\mathbf{x}_d$  (MacEachern, 1999). Let  $\theta_d^{(j)}$  be a draw from the posterior distribution of the parameters that index that model, where  $j$  indicates the iteration in the MCMC algorithm. We set up initial starting values for each source record's  $y_{ic}$  so that the source ranks are preserved. We then update  $y_{ic}$  for each source record sequentially using Gibbs sampling: we sample from the truncated posterior distribution of  $y_{ic}$

given  $\theta_d^{(j)}$  with truncation points defined by the values of  $y_{ic}$  at the  $(i - 1)$ th and  $(i + 1)$ th ranks in the source data. This is shown graphically in Figure 3. This process is repeated until the imputation values settle down into a stable distribution. We repeat this process for other sampled  $\theta_d^{(j)}$  (for a well-spaced sequence of  $j$  values) to get the multiple imputations of  $\mathbf{y}_c$ . It is possible to update the missing  $y_{ic}$  at each iteration of the MCMC; however, we have found that can lead to numerical instabilities.

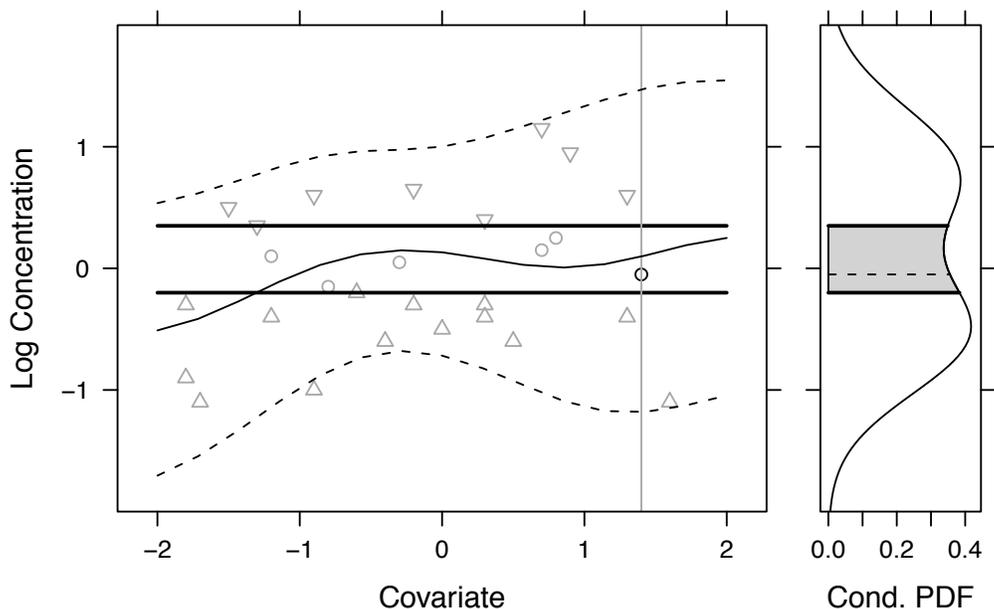


Figure 3: Schematic of the method of rank-preserving predictions (RPP). In the left-hand panel, the curving lines summarize the regression model at a particular iteration in the MCMC, as implied by a single, drawn  $\theta_d^{(j)}$ . The point being updated is the black circle. Gray symbols are current imputed values of the destination lab measurements, with circles for ties in the source lab scale and triangles for observations that must be larger or smaller than the update. Black horizontal lines give the bounds for the update, as dictated by the triangles. The right-hand panel displays the conditional density for the update, with the area of allowable draws in gray.

As a note on practical implementation, when the initial imputed values in the Gibbs sampler are poorly chosen, the Gibbs updates can be slow to mix, especially when the source observations are not in a coarse scale. We have found that making a set of predictions (conditional on a single draw from the posterior of the density regression) that does not respect the ordering forms the basis of a useful starting point for the imputed values. We set the starting quantiles of the imputed values at the empirical quantiles of the draw. It would also be possible to take an annealing approach, starting with a coarse scale where the imputations mix more easily, and gradually enforcing the full observed ordering.

### 3.3 Matched Conditional Quantiles

RPP incorporates covariate information in an auxiliary manner, with the source ranks trumping the covariates. However, in some settings it makes more sense to preserve source rankings within covariate patterns than to preserve them across all source records. For an example in an educational testing context, suppose that questions on an initial version of a test disfavor selected demographic groups—e.g., the content is unfamiliar to them—and that a later version of the test is fair to all groups. A global rank preservation method like RP or RPP would force individuals in the disfavored subgroups to be inaccurately imputed as low scoring on the fair test. It makes more sense to preserve ranks conditional on demographic profile, since one would expect students who score low compared to their like-profiled peers on the unfair test to score low on the neutral test as well.

MCQ is designed to preserve rankings of  $Y$  within covariate patterns. To implement MCQ, we fit two DDP models for  $Y$  given  $X$ : one using the destination-scale observations and the other for the source-scale observations. The models condition on the same covariates, but they are estimated independently. To impute the missing

elements of  $\mathbf{y}_c$ , we draw a value of  $\theta_s^{(j)}$  from the posterior distribution of the parameters in the source DDP model. For each record  $i$  in the source data, we use the drawn  $\theta_s^{(j)}$  to compute the conditional quantile corresponding to the observed  $y_{is}$ ; call this quantile  $q$ . We then draw a value of  $\theta_d^{(j)}$  from the posterior distribution of the parameters in the destination DDP model. We use the drawn  $\theta_d^{(j)}$  to compute the value of the destination scale at the  $q$ th conditional quantile among records with covariate pattern  $x_i$ . This process is displayed graphically in Figure 4. We repeat this process multiple times to get the multiple imputations of  $\mathbf{y}_c$ .

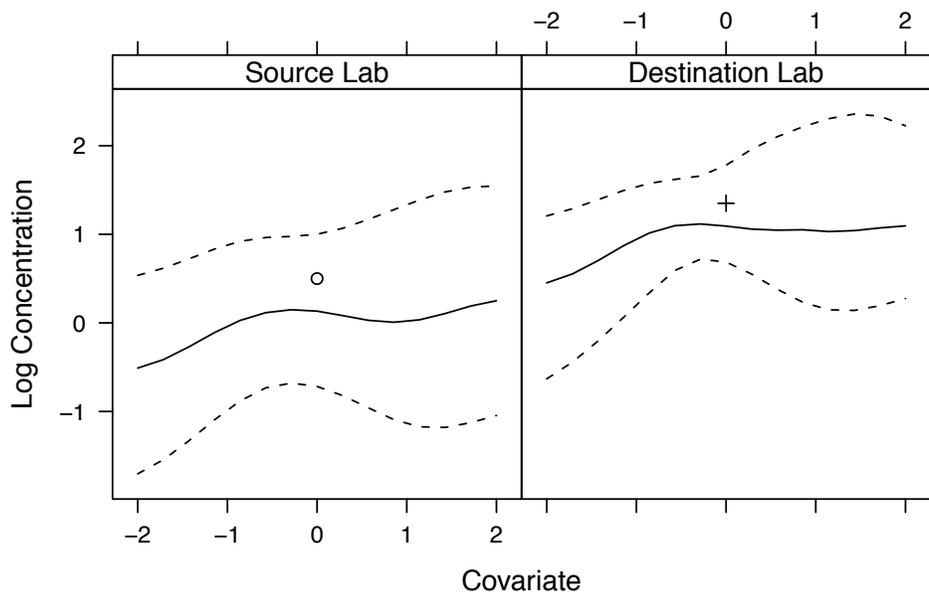


Figure 4: Schematic of matched conditional quantiles (MCQ) approach, with median and 95% predictive bounds for the density regressions, conditional on drawn  $\theta_s^{(j)}$  and  $\theta_d^{(j)}$  values. The observed value  $y_{is} = 0.5$  (circle) is approximately at the  $q = 0.73$  conditional quantile when  $x_i = 0$ . This quantile corresponds to 1.35 in the destination lab regression (plus sign), which becomes the imputed value.

## 4 ILLUSTRATIVE SIMULATIONS

To illustrate the performances of the three methods, we undertake a series of simulation studies. The simulations involve a full factorial design for three binary factors. The first factor is whether or not the covariate matrix  $X$  has a similar distribution in the destination and source data; we call this the balance factor. We expect imbalance in  $X$  to result in comparatively poor performance for RP, whereas RPP and MCQ are intended to adjust for imbalance. The second factor pertains to whether or not there are many ties in the marginal rankings of  $Y$ ; we call this the coarseness factor. Some settings, including the motivating HPHBS example, have ordered categorical data with many ties in at least one of the scales, as opposed to approximately continuous data with few if any ties. Ties can be problematic for the RP method because a small change in the imputed rank can imply a large change in the imputed value. The third factor is whether the transformation function from one scale to the other preserves ranks of  $Y$  globally or only locally. Global preservation of ranks underlies RPP, whereas local preservation of ranks within covariate patterns underlies MCQ.

We generate data from this factorial design using one measurement variable  $Y$  and two covariates  $(X_0, X_1)$ . We set sample sizes  $n_s = 700$  in the source scale and  $n_d = 300$  in the destination scale, which are similar to the sample sizes in the HPHBS application. For any level of the factorial design, we generate replications as follows.

- IF BALANCED: Generate  $X_{i,0} \sim \text{Bern}(.5)$  for all  $i$ .
- IF NOT BALANCED: Generate  $X_{i,0} \sim \text{Bern}(p_i)$ , where  $p_i = 0.25$  for the  $n_s$  source lab observations and  $p_i = 0.75$  for the  $n_d$  destination lab observations.
- Generate  $Y = X_0 + 0.5N(0, I)$ .
- Generate  $X_1 = X_0 + 0.5Y + 0.2N(0, I)$ .

- IF GLOBAL: Transform the source lab observations via the function  $f(y) = -0.5 \exp\{-1 + y\}$ .
- IF LOCAL: Transform the source lab observations via the function  $f(y; x_0) = -0.5 \exp\{-1 + y - x_0\}$ .
- IF COARSE: Round the transformed source lab observations to the nearest 0.5.

We evaluate the abilities of the methods to estimate the regression coefficient of  $Y$  in the regression of  $X_1$  on  $(Y, X_0)$ . Because of the computational demands of the MCMC, we limit the simulation study of RPP and MCQ to ten simulations in each of the eight scenarios. The parameters used to simulate the data are chosen to highlight relative advantages of the methods in various situations so that differences appear even with a small number of simulated repetitions.

Figure 5 summarizes the results of the full factorial simulation study. For comparison, it also includes results from using a method of moments approach to put all source-scale data on a common scale, i.e., we transform the source-scale values to have the same mean and standard deviation as the destination-scale values. In all cases, this simple approach fails to result in unbiased estimates of the regression coefficient. In contrast, the RP method performs favorably when the background covariates are roughly balanced, the source lab scale is not coarse, and the ranks are preserved globally. In these situations, RP performs well even though it ostensibly ignores the strong correlations between  $X$  and  $Y$ . This is because most of the information about the transformed source lab values is contained in the observed ranks, so that preserving ranks essentially preserves correlational structures. In more extensive comparisons, we found that the RP method strongly outperformed the method of moments approach and typically resulted in low bias and proper coverage rates regardless of the correlational structure in the data, provided that the scales are not

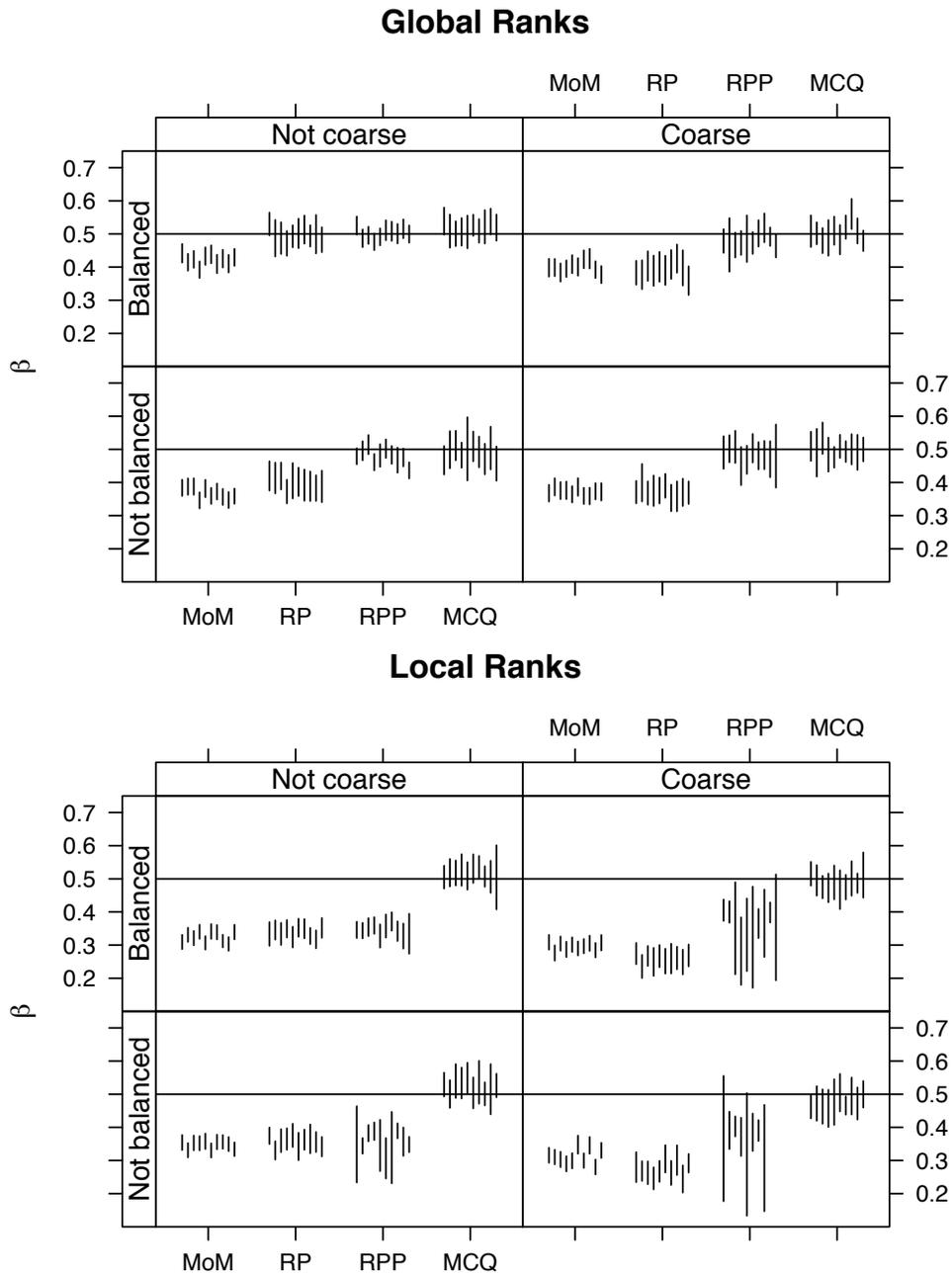


Figure 5: 95% confidence intervals for the regression coefficient associated with the variable measured in two scales. The true value the regression coefficient is 0.5. The simulation examines balanced/unbalanced covariate configurations, coarse/continuous source lab scales, and local/global rank preservations.

coarse, the background covariates are balanced, and global rank preservation holds. However, when any of those three conditions are violated, the performance of the RP method degrades substantially, as evidenced by the large bias in the estimated coefficient.

The RPP method results in approximately unbiased estimates in the four scenarios where global rank preservation holds. The RPP does not suffer from bias due to imbalanced covariates (when global rank preservation holds) because it makes use of background information to anchor imputations. It does not suffer from bias due to coarseness (when global rank preservation holds) because it makes use of covariates to smooth out the coarseness in the source scale measurements. When only local rank preservation holds, RPP results not only in biased estimates, but some of the intervals have large widths. This results from the poor fit of models that incorrectly presume globally rank-preserved predictions, which can yield instability in the imputed values of the observations that are observed in the source lab; it is not a product of inadequate convergence in the MCMC.

The MCQ method is the only method that results in approximately unbiased estimates in all eight scenarios. However, this flexibility comes with a price: the intervals can have comparatively larger widths. For example, in the balanced and not coarse condition with globally-preserved ranks, the confidence intervals resulting from RPP are uniformly narrower than those from MCQ, while still displaying good coverage. Also, if it is the case that the source lab has very few observations, we would expect the source lab model to be quite sensitive to the prior specification.

The results in Figure 5 suggest a two-step decision process for determining which methods can be used, as summarized in Figure 6. First, the analyst should ask whether or not it is sensible to assume global rank preservation. As the stronger assumption, global rank preservation is less flexible than the local assumption, but

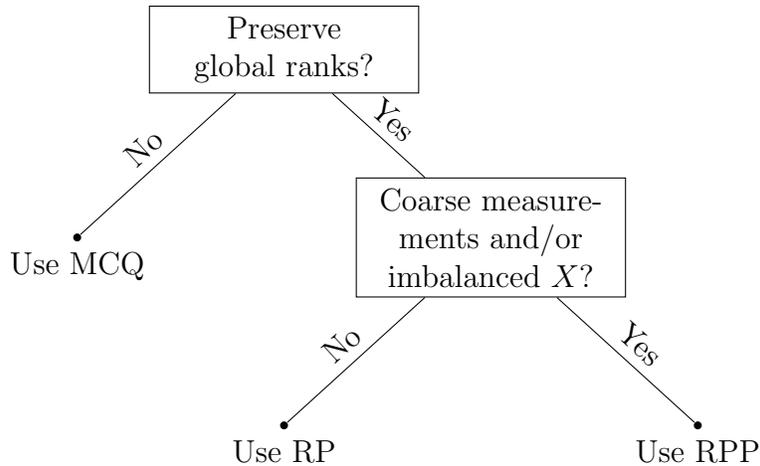


Figure 6: Flow chart summarizing recommended imputation type for various situations.

assuming it results in simpler procedures and possible efficiency gains when global rank preservation is true. Thus, when preserving global ranks is not sensible, or when there is insufficient basis to decide on the local versus global distinction, the analyst should use MCQ; otherwise, the analyst should choose between RPP and RP. When the  $Y$  values are coarse—such that a small change in the imputed rank can correspond to a large change in the imputed  $Y$  value—we recommend RPP. Coarseness in this sense will typically correspond to discrete-valued measurements or multimodality where the modes are well-separated. These can be detected visually in graphs of the marginal distributions of  $Y$ . We also recommend RPP when the distributions of background covariates differ in the two sources. This can be assessed via a regression model of the scale indicator as a function of covariates in  $X$ , much like diagnostics for covariate balance in propensity score matching contexts (Stuart, 2010). When the  $Y$  values are not coarse and the  $X$  values are relatively balanced (and global rank preservation is sensible), the simulations suggest that analysts can use the RP method.

These recommendations also account for the relative computational expenses of the three algorithms. Of the three approaches, the RP method demands the smallest computational burden, requiring only calculations that are essentially instantaneous. Because the resulting draws are independent, the analyst does not need to worry about Markov chain convergence. The other two approaches require density regressions that are more computationally demanding, with the MCQ method calling for two such regressions; this makes the computational load nearly twice as heavy for the MCQ, though not much extra programming effort is required.

## **5 APPLICATION TO ASSAY LAB CHANGES IN THE HPHBS**

We now turn to the mid-study lab assay change in the HPHBS. We focus on measurements of blood lead levels, although some of the other metals also had dissimilar distributions in the two labs. Of the 1435 women, 323 have blood lead levels measured on the destination scale; 807 are measured on the source scale; and, the remainder are missing a lead measurement. Although typically one would rather the destination scale have more observations than the source scale so as to reduce reliance on imputations, the investigators specified the second set of measurements as the destination scale because it offers finer resolution and lower detection limits. We also transform to the log concentration scale so that negative imputations are not a concern.

Based on scientific grounds, we find little reason to believe that one or both of the labs would use a scale that reports different measurements depending on background covariates. Hence, we believe it is sensible to assume global rank preservation when imputing to a common scale. Therefore, we do not use MCQ. As mentioned in Section 2, maternal race is not balanced across laboratory assignments. Additionally,

the source lab observations are coarse, as they are reported in an integer-valued scale (Figure 1). For these reasons, we prefer RPP over RP. As covariates in  $X$ , we include race, age, self-reported smoking status (non-smoker, quit, smoker), and birth weight rounded to the nearest 500g. Exploratory regression analyses indicate that these variables are associated with lead levels.

Because of the modest sample size in the destination scale, the density regression is somewhat sensitive to the prior specifications, especially for the parameter that records the conditional variances of  $y_i$  (called  $\sigma_j^2$  in Appendix A). We judge the suitability of prior specifications based on the resulting marginal distributions of the transformed lab values. In our experience, there is a range of specifications where the marginal distributions are insensitive to the prior distribution, and zones where the marginal distribution of the transformed values is too diffuse or too concentrated to be plausible. The study is still accruing destination-scale data, so that the sensitivity to the prior distribution should diminish as  $n_d$  increases.

The data have missing values for several other variables, although the covariates in the models for RPP are essentially fully observed. We first run the RPP method to form  $m = 10$  completed sets of lead observations in the destination lab scale. As shown in Figure 7, the distributions of the transformed source lab measurements are comparable to the observed destination lab measurements. For each of the completed sets of lab observations, we perform a single imputation for any other missing values via chained equations (Van Buuren and Oudshoorn, 1999; Raghunathan et al., 2001) using a classification and regression tree-based approach (Burgette and Reiter, 2010).

Using the completed datasets, we estimate several quantile regressions (Koenker and Bassett Jr, 1978; Koenker and Hallock, 2001) involving birth weight and mothers' blood lead levels. In this analysis, we restrict our attention to the non-Hispanic black mothers. The models include the baby's gender, an indicator of whether this was the

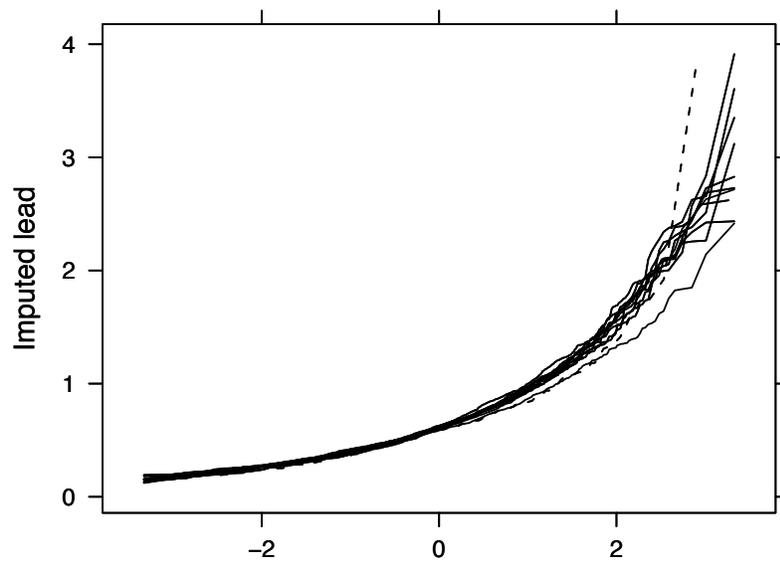


Figure 7: Normal quantile-quantile plots of ten realizations from the RPP method (solid lines) and the observed destination lab observations (broken line).

mother’s first pregnancy, the mother’s age and age squared; all of these are known to be important correlates of birth weight (e.g., Koenker and Hallock, 2001; Abrevaya and Dahl, 2008). The models also include lead, an indicator of whether the mother is a current smoker or not, and their interaction. We include the interaction because exploratory data analyses involving the lead measurements from the source lab suggest it may be important.

Table 1 displays the results of quantile regressions at the 10th through 90th percentiles of birth weight. The lead/smoking interaction is estimated to be negative across the range of response quantiles. For the low response quantiles, 95% confidence intervals for the interaction do not cover zero. These results—including the positive estimates for lead exposure—are similar to those from source lab scale, where the exploratory analysis was performed.

Although the lead/tobacco interaction is the product of high-dimensional exploratory analysis, epidemiological considerations suggest that it deserves attention. Lead exposure has been linked causally to increased blood pressure (Navas-Acien et al., 2007), and nicotine exposure causes short-term spikes in blood pressure (Omvik, 1996). Hypertension is in turn associated with pre-term births (Miranda et al., 2010). On the other hand, smoking during pregnancy surprisingly reduces the risk of preeclampsia (Cnattingius et al., 1997). A primary symptom of preeclampsia is elevated maternal blood pressure, and the condition can be an indication to induce birth. These results suggest that—to improve our understanding of adverse birth outcomes—we should carefully consider the effects of lead exposure, tobacco exposure, hypertension, and their interactions. Such work is part of our ongoing research agenda, and the ability to sensibly aggregate measurements from two laboratories is key to this effort, especially as the study accrues more data in the destination lab scale.

## 6 FINAL REMARKS

We conclude with a brief discussion of applications of the methods described in this article beyond harmonizing laboratory assay data. For instance, the precise wording of census or survey questions may change over time (Jaeger, 1997). It may not be practical to ask individuals multiple versions of the same question, yet longitudinal comparisons may require data on common scales. In large-scale epidemiologic or psycho-social contexts, analysts may seek to combine information from multiple datasets in which key variables are measured or defined differently. Without access to a validation sample on which individuals are measured with the multiple methods, these methods can offer an approach to data harmonization. In education and other contexts, there can be significant rater-to-rater differences (Johnson, 1996). If these differences are not simply additive shifts, it may be desirable to flexibly put all raters' scores on one scale.

## APPENDIX A: GAUSSIAN AND DIRICHLET PROCESSES

Recent Bayesian research has demonstrated the flexibility of mixture modeling approaches (e.g., Escobar and West, 1995; Müller et al., 1996; Griffin and Steel, 2006; Dunson et al., 2007; Dunson and Park, 2008). The Dirichlet process (DP) (Ferguson, 1973; Blackwell and MacQueen, 1973) has become a popular choice for the mixing distribution in such models. Technically, the DP describes a distribution on a collection of distributions that are defined on some measurable space  $\Theta$ . The DP is parametrized by a base measure  $G_0$  defined on  $\Theta$  and a concentration parameter  $\alpha$ , which we will write  $G \sim \text{DP}(\alpha, G_0)$ .

Sethuraman (1994) showed that the DP can be constructed via a stick-breaking process. If  $G \sim \text{DP}(\alpha, G_0)$ , then we can write

$$G = \sum_{j=1}^{\infty} p_j \delta_{\theta_j}, \quad \text{with} \quad \theta_j \stackrel{\text{iid}}{\sim} G_0$$

where  $\mathbf{p} = \{p_j\}$  are drawn according to the so-called stick-breaking construction. If we start with a stick of unit length, and break off a segment of length  $v_1 \sim \text{beta}(1, \alpha)$ , then the first mixture weight  $p_1 = v_1$ . From the portion of the stick that remains, we remove a proportion  $v_2 \sim \text{beta}(1, \alpha)$  of it as the next mixture weight, so  $p_2 = v_2(1 - v_1)$ . This continues on so that in general

$$p_j = v_j \prod_{k=1}^{j-1} (1 - v_k) \quad \text{with} \quad v_k \stackrel{\text{iid}}{\sim} \text{beta}(1, \alpha),$$

which is often written as  $\mathbf{p} \sim \text{GEM}(\alpha)$ . From this definition, one can see that a smaller  $\alpha$  value will typically result in a few heavily-weighted components, with the weights decaying very quickly since  $v_k$  values will be close to one on average. A larger  $\alpha$  will result in mixture weights that decay more slowly.

This constructive representation makes it clear that the DP would be a poor choice for a data model for a continuous response, since it is almost surely discrete. However, as a mixing distribution, this discreteness induces desirable sparsity:  $n$  data points typically will be assigned to fewer than  $n$  mixture components.

The dependent Dirichlet process (DDP) (MacEachern, 1999; De Iorio et al., 2004; Gelfand et al., 2005) induces a DP at each covariate value, but allows for flexible sharing of information across the covariate space. We adopt the DDP that takes on the form

$$G(x) = \sum_{j=1}^{\infty} p_j \delta_{\eta_j(x)}, \quad \text{with} \quad \eta_j \stackrel{\text{iid}}{\sim} G_{0x} \tag{1}$$

where  $\eta_j$  are IID realizations of a base Gaussian process (GP)  $G_{0\mathcal{X}}$  defined on the covariate space  $\mathcal{X}$  (Fronczyk and Kottas, 2010). This is a “single  $p$ ” DDP, as the  $p_j$  values are fixed across the covariate space.

The sharing of information across covariate values is a consequence of the continuity of realizations of the base stochastic process  $G_{0\mathcal{X}}$  (e.g., Rasmussen and Williams, 2006). Conditional on hyperparameters,  $G_{0\mathcal{X}}$  is parametrized so that  $E\eta_j(\mathbf{x}_i) = \mathbf{x}_i'\beta$ ,  $\text{Var}(\eta_j(\mathbf{x}_i)) = \sigma_\eta^2$ ,  $\text{Corr}(\eta_j(\mathbf{x}_i), \eta_j(\mathbf{x}_j)|\phi) = \exp(-\phi|\mathbf{x}_i - \mathbf{x}_j|^2)$  with  $\phi > 0$  for any  $\mathbf{x}_i, \mathbf{x}_j \in \mathcal{X}$ . We collect these parameters as  $\boldsymbol{\psi} = (\beta, \sigma_\eta^2, \phi)$ .

Our hierarchical model is then

$$y_i \stackrel{\text{iid}}{\sim} N(\eta_{w(i)}(\mathbf{x}_i), \sigma_{w(i)}^2) \quad (2)$$

$$\Pr(w(i) = j) = p_j \quad (3)$$

$$\mathbf{p} \sim \text{GEM}(\alpha) \quad (4)$$

$$\eta_j(\cdot) \stackrel{\text{iid}}{\sim} G_{0\mathcal{X}}(\cdot; \boldsymbol{\psi}) \quad (5)$$

$$\sigma_j^2 \stackrel{\text{iid}}{\sim} \text{inv-gamma}(a_\sigma, b_\sigma) \quad (6)$$

$$\alpha \sim \text{gamma}(a_\alpha, b_\alpha) \quad (7)$$

$$\phi \sim \text{unif}(0, b_\phi) \quad (8)$$

$$\beta \sim \text{normal}(0, B_0^{-1}) \quad (9)$$

$$\sigma_\eta^2 \sim \text{inv-gamma}(a_\eta, b_\eta) \quad (10)$$

In practice, we choose to truncate the DP such that the stick-breaking representation of  $G \sim \text{DP}(\alpha, G_0)$  is

$$G = \sum_{j=1}^L p_j \delta_{\theta_j} \quad (11)$$

by assigning  $p_L = 1 - \sum_{k=1}^{L-1} p_k$  for a fixed  $L$ . This allows us to use the blocked Gibbs

sampler of Ishwaran and James (2002), which samples the mixture components  $w(i)$  jointly. (See also Ishwaran and James (2001).) Otherwise, it is possible to use the full DP and sample according to the Polya urn representation conditioning on the other the others. See the Appendix B for details of the MCMC algorithm.

If observations  $y_i$  are rounded to a small number of possible outcome values, or if there is a known detection limit associated with the measurement, then the conditional normality implied by our model may be unrealistic. In such cases we augment the model with latent quantities that represent the pre-rounding quantity, or the quantity that was not truncated at the detection limit. This standard data augmentation method is easy to add to the proposed model (Tanner and Wong, 1987).

## APPENDIX B: MCMC DETAILS

Following Rasmussen and Williams (2006), we use  $K(X_1, X_2)$  to denote matrix of pairwise GP covariances (conditional on the mixture indicator) between the points described by the rows of  $X_1$  and  $X_2$ . We factor  $K(X_1, X_2) = \sigma_\eta^2 H(\phi)$ . Further, we denote with  $X_u$  the matrix of unique predictor values.

Updates should be as follows:

- Update  $\eta_j$  for  $j = 1, \dots, L$ .
  - If no observations are currently assigned to the  $j$ th mixture component, then  $\eta_j \sim N(X_u\beta, K(X_u, X_u))$ .
  - Else,  $\eta_j | \text{all} \sim \text{normal}(\mu_\eta, \Sigma_\eta)$  where

$$\begin{aligned} \mu_\eta &= X_u\beta + K(X_u, X_j)[K(X_j, X_j) + \sigma_j^2 I]^{-1}(y_j - X_j\beta) \\ \Sigma_\eta &= K(X_u, X_u) - K(X_u, X_j)[K(X_j, X_j) + \sigma_j^2 I]^{-1}K(X_j, X_u). \end{aligned}$$

Here,  $X_j$  and  $y_j$  collect the observations that are assigned to the  $j$ th mixture component.

(See Chapter 2 of Rasmussen and Williams (2006).)

- For  $j = 1, \dots, L$ , update

$$\sigma_j^2 \sim \text{inv-gamma}(a_\sigma + .5n_j^*, b_\sigma + .5 \sum_{i:w(i)=j} (y_i - \eta_j(\mathbf{x}_i))^2)$$

where  $n_j^*$  counts the number of elements assigned to the  $j$ th mixture component.

- For  $i = 1, \dots, n$ , sample

$$w(i) \sim \sum_{j=1}^L \tilde{p}_j \delta_j(w(i)) \text{ where } \tilde{p}_j \propto p_j N(y_i; \eta_j(\mathbf{x}_i), \sigma_j^2)$$

- Update  $\mathbf{p}$ :

$$\mathbf{p} \sim \text{generalized-Dir}((n_1^*, \dots, n_{L-1}^* + 1); (\alpha + \sum_{j=2}^L n_k^*, \dots, \alpha + n_L^*)),$$

which can be sampled as described in Ishwaran and James (2002) or Fronczyk and Kottas (2010).

- Sample  $\alpha$ :

$$\alpha \sim \text{gamma}(\text{shape} = L + a_\alpha, \text{rate} = b_\alpha - \log(p_L))$$

- Sample  $\beta \sim \text{normal}(\hat{\beta}, B^{-1})$  where

$$B = n_\eta^* X_u' K^{-1} X_u + B_0 \quad \text{and} \quad \hat{\beta} = B^{-1} \sum_{j:n_j^* > 0} X_u K^{-1} \eta_j(X_u).$$

Here,  $K$  is shorthand for  $K(X_u, X_u)$ , and  $n_\eta^*$  counts the number of mixture components that have at least one assigned observation.

- Following Fronczyk and Kottas (2010), we specify the prior  $p(\phi) \propto \mathbf{1}\{\phi < b_\phi\}$ . We also require  $\phi > a_\phi$  for a small  $a_\phi$  to avoid proposing  $H(\phi)$  matrices that are ill-conditioned. We sample  $\phi$  in a random walk MH step with the conditional density proportional to

$$|H(\phi)|^{-n_\eta^*/2} \exp\left(-.5\sigma_\eta^{-2} \sum_{j:n_j^*>0} (\eta_j - X_u\beta)' H^{-1}(\phi)(\eta_j - X_u\beta)\right) \mathbf{1}\{a_\phi < \phi < b_\phi\}.$$

- Sample  $\sigma_\eta^2$  from

$$\text{inv-gamma}(a_\eta + .5n_\eta^*, b_\eta + .5 \sum_{j:n_j^*>0} (\eta_j - X_u\beta)' H^{-1}(\phi)(\eta_j - X_u\beta)).$$

To generate  $m$  completed datasets, record  $m$  well-spaced draws of the parameters from the density regression described above. For each of these  $m$  draws, sample imputed values in the destination lab scale as follows:

- For RPP, generate a starting set of imputed values. Then repeatedly update the imputed values one at a time. Let  $b_L$  be the maximum of the current imputed values that are required to be smaller than the observation whose value we are updating. Let  $b_U$  be the minimum of the observations that are required to be larger. We then wish to update the imputed value from the conditional density, restricted to be in the interval  $(b_L, b_U)$ . To achieve this, first calculate the probability that the restricted draw will come from the  $j$ th mixture component, which is proportional to  $p_j[\Phi(b_U; \eta_j(\mathbf{x}_i), \sigma_j^2) - \Phi(b_L; \eta_j(\mathbf{x}_i), \sigma_j^2)]$ , where  $\Phi$  is the normal CDF. After sampling the mixture indicator, sample the imputed value

according to a truncated univariate normal distribution. Observations without a measurement in either lab are easily handled: simply impute into the destination lab scale without any truncation in the conditional distributions.

- For MCQ, determine the conditional quantile for each source lab observation in the source lab scale, which is described by a linear combination of normal CDF values. Numerically invert the conditional CDFs in the destination lab scale to produce the imputed values.

## References

- Abrevaya, J., and Dahl, C. (2008), “The effects of birth inputs on birthweight,” *Journal of Business and Economic Statistics*, 26(4), 379–397.
- Blackwell, D., and MacQueen, J. (1973), “Ferguson distributions via Pólya urn schemes,” *The Annals of Statistics*, 1(2), 353–355.
- Burgette, L. F., and Reiter, J. P. (2010), “Multiple imputation for missing data via sequential regression trees,” *American Journal of Epidemiology*, 172(9), 1070–1076.
- Cnattingius, S., Mills, J., Yuen, J., Eriksson, O., and Salonen, H. (1997), “The paradoxical effect of smoking in preeclamptic pregnancies: Smoking reduces the incidence but increases the rates of perinatal mortality, abruptio placentae, and intrauterine growth restriction,” *American Journal of Obstetrics and Gynecology*, 177(1), 156–161.
- Cole, S. R., Chu, H., and Greenland, S. (2006), “Multiple-imputation for measurement-error correction,” *International Journal of Epidemiology*, 35, 1074–1081.

- De Iorio, M., Mueller, P., Rosner, G., and MacEachern, S. (2004), “An ANOVA model for dependent random measures,” *Journal of the American Statistical Association*, 99(465), 205–215.
- Dunson, D., and Park, J. (2008), “Kernel stick-breaking processes,” *Biometrika*, 95(2), 307.
- Dunson, D., Pillai, N., and Park, J. (2007), “Bayesian density regression,” *Journal of the Royal Statistical Society: Series B*, 69(2), 163–183.
- Durrant, G. B., and Skinner, C. (2006), “Using missing data methods to correct for measurement error in a distribution function,” *Survey Methodology*, 32, 25–36.
- Escobar, M., and West, M. (1995), “Bayesian density estimation and inference using mixtures,” *Journal of the American Statistical Association*, 90(430).
- Ferguson, T. (1973), “A Bayesian analysis of some nonparametric problems,” *The Annals of Statistics*, 1(2), 209–230.
- Fronczyk, K., and Kottas, A. (2010), “A Bayesian nonparametric modeling framework for developmental toxicity studies,” *University of California, Santa Cruz Technical Report*, UCSC-SOE-10(11), 1–34.
- Gelfand, A., Kottas, A., and MacEachern, S. (2005), “Bayesian nonparametric spatial modeling with Dirichlet process mixing,” *Journal of the American Statistical Association*, 100(471), 1021–1035.
- Griffin, J., and Steel, M. (2006), “Order-based dependent Dirichlet processes,” *Journal of the American Statistical Association*, 101(473), 179–194.
- Gulson, B., Pounds, J., Mushak, P., Thomas, B., Gray, B., and Korsch, M. (1999), “Estimation of cumulative lead releases (lead flux) from the maternal skeleton

- during pregnancy and lactation,” *The Journal of Laboratory and Clinical Medicine*, 134(6), 631–640.
- Ishwaran, H., and James, L. (2001), “Gibbs sampling methods for stick-breaking priors,” *Journal of the American Statistical Association*, 96(453), 161–173.
- Ishwaran, H., and James, L. (2002), “Approximate Dirichlet process computing in finite normal mixtures,” *Journal of Computational and Graphical Statistics*, 11(3), 508–532.
- Jacobs, D., and Nevin, R. (2006), “Validation of a 20-year forecast of US childhood lead poisoning: Updated prospects for 2010,” *Environmental Research*, 102(3), 352–364.
- Jaeger, D. (1997), “Reconciling the old and new census bureau education questions: Recommendations for researchers,” *Journal of Business and Economic Statistics*, 15(3), 300–309.
- Johnson, V. (1996), “On Bayesian analysis of multirater ordinal data: An application to automated essay grading,” *Journal of the American Statistical Association*, 91(433), 42–51.
- Koenker, R., and Bassett Jr, G. (1978), “Regression quantiles,” *Econometrica*, 46(1), 33–50.
- Koenker, R., and Hallock, K. (2001), “Quantile regression,” *Journal of Economic Perspectives*, 15(4), 143–156.
- MacEachern, S. (1999), Dependent nonparametric processes,, in *Proceedings of the Section on Bayesian Statistical Science*, pp. 50–55.

- Miranda, M., Swamy, G., Edwards, S., Maxson, P., Gelfand, A., and James, S. (2010), “Disparities in maternal hypertension and pregnancy outcomes: Evidence from North Carolina, 1994–2003,” *Public Health Reports*, 125(4), 579.
- Müller, P., Erkanli, A., and West, M. (1996), “Bayesian curve fitting using multivariate normal mixtures,” *Biometrika*, 83(1), 67.
- Navas-Acien, A., Guallar, E., Silbergeld, E., and Rothenberg, S. (2007), “Lead exposure and cardiovascular disease—A systematic review,” *Environmental Health Perspectives*, 115(3), 472.
- Omvik, P. (1996), “How smoking affects blood pressure,” *Blood Pressure*, 5(2), 71.
- R Development Core Team (2010), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.  
**URL:** <http://www.R-project.org>
- Raghunathan, T., Lepkowski, J., Van Hoewyk, J., and Solenberger, P. (2001), “A multivariate technique for multiply imputing missing values using a sequence of regression models,” *Survey methodology*, 27(1), 85–96.
- Rasmussen, C., and Williams, C. (2006), *Gaussian Processes for Machine Learning*, Cambridge, MA: MIT Press.
- Reiter, J. P., and Raghunathan, T. E. (2007), “The multiple adaptations of multiple imputation,” *Journal of the American Statistical Association*, 102, 1462–1471.
- Rubin, D. (1987), *Multiple Imputation for Nonresponse in Surveys*, New York, NY: John Wiley.

- Schenker, N., and Parker, J. D. (2003), “From single-race reporting to multiple-race reporting: Using imputation methods to bridge the transition,” *Statistics in Medicine*, 22, 1571–1587.
- Sethuraman, J. (1994), “A constructive definition of Dirichlet priors,” *Statistica Sinica*, 4(2), 639–650.
- Stuart, E. (2010), “Matching methods for causal inference: A review and a look forward,” *Statistical Science*, 25(1), 1.
- Tanner, M., and Wong, W. (1987), “The calculation of posterior distributions by data augmentation,” *Journal of the American Statistical Association*, 82(398), 528–540.
- Thomas, N., Raghunathan, T. E., Schenker, N., Katzoff, M. J., and Johnson, C. L. (2006), “An evaluation of matrix sampling methods using data from the National Health and Nutrition Examination Survey,” *Survey Methodology*, 32, 217–232.
- Thomas, V. (1995), “The elimination of lead in gasoline,” *Annual Review of Energy and the Environment*, 20(1), 301–324.
- Van Buuren, S., and Oudshoorn, K. (1999), “Flexible multivariate imputation by MICE,” *TNO Prevention Center*, pp. 6–20.
- Venables, W., and Ripley, B. (2002), *Modern Applied Statistics with S*, New York, NY: Springer Verlag.

Percentile	Intercept	Age	Age <sup>2</sup>	Male	First Preg	Lead	Tobacco	Lead/Tobacco
10th	2016 (1684, 2349)	-3 (-26, 20)	-3 (-6, 0)	5 (-184, 193)	26 (-175, 227)	508 (114, 903)	575 (127, 1023)	-828 (-1419, -237)
20th	2502 (2312, 2691)	-2 (-15, 11)	-3 (-5, -1)	61 (-51, 173)	-36 (-150, 79)	321 (111, 531)	237 (-79, 553)	-548 (-983, -113)
30th	2671 (2512, 2829)	-1 (-12, 10)	-2 (-3, 0)	120 (30, 211)	11 (-90, 111)	217 (-15, 448)	135 (-196, 465)	-389 (-829, 51)
40th	2817 (2655, 2979)	-5 (-15, 6)	-1 (-2, 1)	130 (45, 215)	-52 (-146, 43)	231 (-2, 464)	147 (-205, 499)	-389 (-866, 87)
50th	2923 (2773, 3073)	-5 (-15, 5)	0 (-2, 1)	124 (47, 201)	-43 (-131, 46)	252 (55, 448)	93 (-185, 371)	-354 (-736, 28)
60th	3081 (2953, 3210)	-7 (-16, 2)	0 (-1, 1)	121 (44, 198)	-48 (-131, 35)	223 (43, 402)	-32 (-262, 198)	-272 (-583, 39)
70th	3209 (3063, 3356)	-7 (-16, 3)	0 (-1, 1)	124 (46, 201)	-56 (-144, 32)	221 (48, 393)	-45 (-277, 187)	-236 (-528, 57)
80th	3439 (3303, 3574)	-5 (-14, 5)	0 (-1, 0)	85 (1, 168)	-65 (-154, 24)	162 (-9, 333)	-66 (-309, 178)	-178 (-448, 92)
90th	3660 (3489, 3830)	-1 (-13, 12)	-1 (-2, 0)	89 (-14, 192)	-113 (-227, 0)	234 (12, 455)	-14 (-482, 454)	-284 (-880, 311)

Table 1: Point estimates and 95% confidence intervals for quantile regressions suggested by exploratory analyses performed using source lab lead observations. The coefficients in this table refer to the destination lab lead scale. “Age” is centered maternal age; “First Preg” is a dummy variable equaling one if it is the mother’s first pregnancy; and “Tobacco” is a dummy for self-reported smoker/non-smoker status during the pregnancy.