Bayesian Analysis (2007) **2**, Number 1, pp. 1–30

Does the effect of micronutrient supplementation on neonatal survival vary with respect to the percentiles of the birth weight distribution?

Francesca Dominici*, Scott L. Zeger†, Giovanni Parmigiani‡, Joanne Katz§, and Parul Christian¶,

Abstract.

Scientific Background: In developing countries, higher infant mortality is partially caused by poor maternal and fetal nutrition. Clinical trials of micronutrient supplementation are aimed at reducing the risk of infant mortality by increasing birth weight. Because infant mortality is greatest among the low birth weight infants (LBW) ($\leq$ 2500 grams), an effective intervention may need to increase birth weight among the smallest babies. Although it has been demonstrated that supplementation increases the birth weight in a trial conducted in Nepal, there is inconclusive evidence that the supplementation improves their survival. It has been hypothesized that a potential benefit of the treatment on survival among the LBW infants is partly compensated by a null or even harmful effect among the largest infants. Exploratory analyses have suggested that the treatment effect on birth weight might vary with respect to the percentiles of the birth weight distribution.

Data: The methods in this paper are motivated by a double-blind randomized community trial in rural Nepal (Christian et al 2003a,b). The investigators implemented an intervention program to evaluate benefits of the following micronutrient supplementations: folic acid and vitamin A (F+A); folic acid, iron, and vitamin A (F+I+A); folic acid, iron, zinc, and vitamin A (F+I+Z+A); multiple nutrients and vitamin A (M+A). Each micronutrient supplement was administered daily to 1000 pregnant women, who ultimately delivered approximately 800 live-born infants. The team measured the birth weight within 72 hours of delivery and then followed the infants for one year to determine whether or not they survived. In addition, they measured several characteristics of the mother (maternal age, maternal height, arm circumference) and of the infant (weight, length, head and chest circumference).

*Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, http://www.biostat.jhsph.edu/~fdominic/

†Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, http://www.biostat.jhsph.edu/~szeger/

‡The Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University, Baltimore, MD, http://astor.som.jhmi.edu/~gp/

§Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, http://faculty.jhsph.edu/?F=Joanne&L=Katz

¶Department of Internationa Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, http://faculty.jhsph.edu/?F=Parul&L=Christian

In this case study we focus on the supplementations F+I+A and M+A as compared to vitamin A only and we address the following scientific questions:

1. Is there an overall effect of the treatments on birth weight? Does this effect vary with the percentiles of the birth weight distribution? In particular, is it largest among the LBW infants?
2. Is there an overall effect of the treatments on survival? Does this effect vary with the percentiles of the birth weight distribution? In particular, is it largest among the LBW infants?
3. Do these percentile-specific effects on birth weight and survival differ by micronutrients?

Statistical Approach: The data analysis is challenged by measurement error and informative missing data in birth weight and survival. In community-based interventions in developing countries, most births occur in the home without assistance from trained birth attendants. Approximately 88% of the babies are measured within 72 hours of the delivery. The remaining 12% are measured between 72 and 2000 hours from the delivery approximately. Hence, weights are obtained at varying times following birth and therefore are imprecise measures of the "true weight at birth". In addition, a high proportion of deaths of young infants occur in the first few hours after birth. If there is a delay in reaching the mother and infant, then many of these infants would not be weighed because they have already died. For example in the F+I+A group, approximately 7% of the birth weight measurements are missing and among this 7%, approximately 34% of the babies have died within 24 hours of the delivery. These babies are likely to have been of lower birth weight than those who survived to be weighed, and therefore, these missing birth weights due to death are likely to be informative.

In this paper we develop a measurement error model with counterfactual variables that address the scientific questions for this birth weight-mortality case study. Our approach integrates Bayesian methods and data augmentation (Tanner and Wong 1987; Tanner 1991; Albert and Chib 1993; Chib and Greenberg 1998) with a counterfactual model and principal stratification (Rubin 1978; Holland 1986; Frangakis and Rubin 2002). We calculate marginal posterior distributions of the treatment effects on birth weight and infant mortality that are allowed to vary with the percentiles of the birth weight distributions. We compare our posterior inferences with two simpler approaches. The first still relies on a Bayesian approach but ignores the uncertainty in the imputation and prediction of the birth weight and does account for the mother's covariates. The second is a simpler re-sampling approach that imputes the missing birth weights (Rubin 1987).

Results and Public Health Impact: First we found that both F+I+A and M+A increase birth weight. However, the F+I+A increases birth weight mainly among the LBW infants, whereas M+A increases birth weight across the entire birth weight distribution compared to vitamin A only. The F+I+A reduces the risk of infant mortality, whereas the M+A slightly increases the risk of early infant mortality, especially among the larger infants.

Currently, recommendations exist to supplement pregnant women in developing countries. This case study provides critical information toward the evaluation and

planning of these public health interventions.

Keywords: percentiles, counterfactual, Bayesian computation

1 Introduction

In developing countries, higher infant mortality is partially caused by poor maternal and fetal nutrition. Because infant mortality is greatest among low birth weight (LBW $\leq$ 2500 grams) and very low birth weight (VLBW $\leq$ 1500 grams) infants, it is assumed that an effective intervention must increase birth weight among the smallest babies, that is, in the left tail of the birth weight distribution. That maternal nutritional supplementation increases average birth weight has been demonstrated in replicated randomized trials in several countries (Lechtig et al. 1975; Ceesay et al. 1997; Caulfield et al. 1999; Christian et al. 2003a). However, to date, there is limited direct evidence that maternal supplementation causes a reduction in the prevalence of babies born at the smallest weights and that this reduction improves their survival (Garner et al. 1992; McIntire et al. 2001; West et al. 1999; Katz et al. 2000a; Rasmussen 2001; Christian et al. 2003b).

The methods in this paper are motivated by a double-blind randomized community trial in rural Nepal (Christian et al. 2003a). The investigators administered an intervention program to evaluate benefits of the following micronutrient supplementations: folic acid and vitamin A; folic acid, iron, and vitamin A; folic acid, iron, zinc, and vitamin A; multiple nutrients and vitamin A. The control was vitamin A alone. Each micronutrient supplement was administered daily to 1000 pregnant women, who ultimately delivered approximately 800 live born infants. Details on the study designs are in Christian et al. (2003a). The team measured the birth weight within 72 hours of delivery and then followed the infants for one year to determine whether or not they survived. In addition, they measured several characteristics of the mother (maternal age, parity, maternal height, arm circumference) and of the infant (weight, length, head and chest circumference).

To develop the methodology, we will focus our data analysis on two novel treatment groups, the folic acid, iron, and vitamin A (denoted as F+I+A) and the multiple nutrient and vitamin A (denoted as M+A), in comparison to the standard control (vitamin A only). The data analysis is challenged by measurement error and informative missing data. In community-based interventions in developing countries, a large proportion of births occur in the home without assistance from trained birth attendants. For example, in the F+I+A group, 88% of the babies were measured within 72 hours of the delivery. The remaining 12% were measured between 72 and 2644 hours of delivery. Hence, the observed weights are imprecise measures of the "birth weight" which we define here as the value at 72 hours.

In addition, a non-negligible proportion of infants die in the first few hours of birth. If there is a delay in reaching the mother and infant, then many of these infants cannot be weighed because they have already died. In the F+I+A group, approximately 7% of

the birth weight measurements are missing and among this 7%, 34% of the babies have died right after the delivery. These babies are likely to have been of lower birth weight than those who survived to be weighed, and therefore, these missing birth weights are likely to be informative of birth weight. Table 2 provides summary statistics for all treatment groups. Gestational age, number of cigarettes smoked, height, weight and age of the mother are all good predictors of birth weight and will be used to impute missing weights.

An interesting aspect of this study is that the investigators anticipate that some of these micronutrient supplementations may affect birth weight, and ultimately survival, differently among the smaller and larger babies. The top panel of Figure 1 shows the difference between the empirical quantile functions of the birth weights for the two novel interventions, each versus the control ($\widehat{Q}_1(p) - \widehat{Q}_0(p)$) plotted against the percentiles p. The red dots denote quantile differences of birth weights including the ones measured after 72 hours. The black dots denote quantile differences obtained from a "working data set" where the birth weight measurements taken after the 72 hours where replaced by their predicted values at time zero (details on this prediction model are provided in Section 2). The dotted horizontal line is placed at the average difference of the birth weights between the two groups. Note that although the average treatment effects for the two treatment groups are similar and equal to 67 and 81 grams for the F+I+A and M+A groups respectively, these plots suggest that there could be an interaction between the treatment effect and the birth weight percentiles: F+I+A increases birth weight mainly among the smaller babies, while the M+A increases birth weight across the entire birth weight distribution.

To explore the association between birth weight and mortality, we fit a logistic regression model expressing the log odds of infant death as a separate smooth function of the birth weight for the control and intervention groups. The bottom panel of Figure 1 shows the estimated smooth curves with 95% confidence bands across the ranges of the measured birth weights in the two groups. These plots suggest that the probability of death decreases as the birth weight increases and tends to rise again for the heaviest babies in the control group.

This exploratory analysis suggest that: 1) the treatment effect on birth weight might vary with respect to the percentiles of the birth weight distribution for F+I+A but not for M+A; 2) the increase in birth weights among the largest babies for M+A could have a negative impact on survival; 3) it is necessary to properly account for the measurement error in the time of the birth weight measurements.

In this paper, we develop a Bayesian measurement error model to address the following scientific questions:

1. Is there an overall effect of the treatments on birth weight? Does this effect vary with respect to the percentiles of the birth weight distribution? In particular, is it largest among the LBW infants?

2. Is there an overall effect of the treatments on survival? Does this effect vary with respect to the percentiles of the birth weight distribution? In particular, is it

largest among the LBW infants?

3. Do these percentile-specific effects on birth weight and survival differ by micronutrients?

The broad objectives of this paper are to address these scientific questions by developing and applying a Bayesian model with counterfactual variables (Rubin 1978; Holland 1986) for this birth weight-mortality study. Our approach integrates Bayesian methods and data augmentation (Tanner and Wong 1987; Tanner 1991; Albert and Chib 1993; Chib and Greenberg 1998) with a counterfactual model with principal stratification (Rubin 1978; Holland 1986; Frangakis and Rubin 2002). We define parameters that measure the effects of an intervention on a clinical outcome (infant mortality) that are allowed to vary with the percentiles of the post-treatment variable (birth weight). A Bayesian approach to counterfactual modelling is very attractive because we can: 1) calculate the posterior distributions of percentile-specific effects accounting for the uncertainty about the missing counterfactuals, measurement error, and missing data; and 2) investigate the sensitivity of causal inferences to key assumptions for which there are no direct observations in the data set.

In our previous work (Dominici et al. 2005b) we have estimated percentile-specific effects for this case study by comparing F+I+A versus A and by using a "working data set" where: a) the missing birth weight measurements were imputed by use of a regression model having as predictors the mother's covariates; and b) the birth weight measurements made after the 72 hours where replaced by their predicted values at time zero. We did not account for the uncertainty in the imputation and prediction, and we relied upon this working data set to make inferences on the parameters of interest.

In this manuscript we extend our previous approach and build a Bayesian measurement error model that: 1) imputes the missing birth weights accounting for the mother's covariates and death; 2) accounts for the uncertainty in the imputation of the missing birth weights and in the prediction of the "weights at birth" for the babies that have been weighted after 72 hours; 3) compares our Bayesian inferences with our previous work (Dominici et al. 2005b) but does not consider the mother's covariates and the uncertainty in the imputation of the birth weights; 4) compares our Bayesian inferences with a non-parametric approach which is based upon smoothing across percentile differences between the empirical quantile functions of the two groups and which "fills in" the missing data by multiple imputation (Rubin 1987); and finally 5) contrast results between the two treatment groups.

2 Details on the community intervention trial

The randomized trial design, methods and results have been described previously (Christian et al. 2003b; Katz et al. 2005). Briefly, 426 communities in the Sarlahi district, Nepal, were randomized to receive one of five different maternal supplements. From December 1998 through April 2001, all married women of childbearing age who were not already pregnant or breastfeeding an infant less than nine months of age and

who agreed to participate, were visited every five weeks and asked if they had experienced menses in the past five weeks. If they had not, they were given a urine-based pregnancy test. If found to be pregnant, they were enrolled in the trial and supplemented with either vitamin A alone as the control group (1000 μ_g), vitamin A plus folic acid (400 μ_g), vitamin A plus folic acid plus iron (60 mg ferrous fumarate), vitamin A plus folic acid plus iron plus zinc (30 mg zinc sulphate), or a multiple micronutrient supplement that included the same quantities of vitamin A, iron folic acid and zinc, along with vitamin D (10 μ_g), vitamin E (10 mg) vitamin B-1 (1.6 mg), vitamin B-2 (1.8 mg), niacin (20 mg), vitamin B-6 (2.2 mg), vitamin B-12 (2.6 μ_g), vitamin C (100 mg), vitamin K (65 μ_g), copper (2.0 mg), and magnesium (100 mg).

Pregnant women were interviewed at the time of enrollment when maternal height, weight, age, date of last menstrual period, parity, smoking history, and other characteristics were recorded. The main outcomes of the study were birth weight and infant survival. Since 95% of births occurred in the home, attended primarily by relatives or untrained traditional birth attendants, a female staff member who lived in the village reported the birth to a supervisor who dispatched an anthropometrist to the home to obtain "birth weight" using a balance scale accurate at $\pm$ 2 grams so that pure measurement error is negligible. The aim was to weigh the infant as soon after birth as possible. The inability to obtain weights at the exact time of birth leads to a set of methodological issues, some of which can be addressed by altering data collection procedures and some of which can be addressed at the time of data analysis. The question is how to use the observed weights and covariates predictive of birth weight to estimate what the birth weight would have been if it had been measured at the time of delivery.

The second issue is that a high proportion of deaths of young infants occur in the first few hours after birth. If there is a delay in reaching the mother and infant, then many of these infants cannot be weighed because they have already died. It is also more likely that these early deaths involve premature and small for gestational age babies. Hence, these missing birth weights due to death are likely to be lower than those of infants who survive long enough for a weight to be obtained. Again, it may be possible to predict the birth weight of these infants through the use of maternal covariates and weights of infants who died soon after birth, but for whom birth weight was obtained.

In this paper we will focus on two treatments only: 1) folic acid plus iron plus vitamin A (which we will denote by F+I+A); and 2) the multiple micronutrient supplement plus vitamin A (which we will denote by M+A). Table 1 summarizes the sample sizes, the percentages of the birth weight measurements made after 72 hours, the percentage of missing birth weights, and the percentages of deaths among the babies with missing birth weight measurements.

3 A Non-parametric approach with multiple imputation

We start the analysis using a simple non-parametric approach with multiple imputation to estimate percentile-specific treatment effects on birth weight. In the results section (Section 5), we will compare results from the approach described here versus a Bayesian model with measurement error and counterfactual variables described in Section 4.

Notation: To establish notation, let $W_{it_i}^{obs}$ be the weight of the infant i measured at time t_i, let Y_i^{obs} be the observed mortality indicator within one year, let Z_i be the treatment indicator, and let $\boldsymbol{x}_i$ be the vector of mother's covariates. Let $I = \{i : i = 1, \ldots, N\}$ be the entire population of babies. We denote by n_0 and n_1 the number of live births for the control and the treatment groups respectively and let $N = n_0 + n_1$ be the total number of live births. The data analysis is challenged by two facts: 1) for $i \in I_{mis} \subset I$, $W_{it_i}^{obs}$ are missing values; 2) for $i \in M \subset I$, $W_{it_i}^{obs}$ are measured for $t_i > 72$ hours. Table 2 summarizes the percentages of missing data and of measurements made after 72 hours for each treatment group.

Multiple imputation of missing birth weights and prediction of "weights at birth": To impute the missing birth weights and predict the birth weights for the babies that have been measured after 72 hours, we fitted the following regression model separately for the two treatment groups compared to the control (that is for F+I+A versus A, and for M+A versus A):

$$\begin{aligned} W_{it_i}^{obs} \mid t_i, Z_i, Y_i^{obs}, \boldsymbol{x}_i &\sim N(\mu_i, \sigma^2), \text{ where} \\ \mu_i &= \beta_0 + \beta_1 Z_i + \beta_2 t_i + \beta_3 Y_i^{obs} + \beta_4 \text{num.cig}_i + \beta_5 \text{gest.age}_i + \\ &\quad + \beta_6 \text{mom.weight}_i + \beta_7 \text{mom.height}_i + \beta_8 \text{mom.age}_i, \\ &\quad i \in I \backslash I_{mis}. \end{aligned} \tag{1}$$

Missing birth weights were multiply imputed by using multiple imputation (Rubin 1987). Specifically, let $\widehat{W}_{it_i}$ be the predicted birth weight at time t_i from model (1). Let $\widehat{\sigma}^2$ be the estimated residual variance of the regression model. For $i \in I_{mis}$, we created fifty imputed data sets by sampling $W_{it_i=0}^{(j)}$ from a normal distribution with mean $\widehat{W}_{it_i=0}$ and standard deviation $\widehat{\sigma}$ for $j = 1, \ldots, J$. For $i \in M$, we predict the"birth weights" by taking $\widehat{W}_{it_i=0} + (W_{it_i}^{obs} - \widehat{W}_{it_i})$. Note that this approach accounts for the uncertainty in the imputation of the missing data but not for the uncertainty in the prediction of the birth weights for the infants measured after 72 hours.

Estimating percentile-specific effects: The second component of this analysis approach is to estimate the treatment effect on birth weight as a smooth function of the percentiles of the birth weight distribution. In this approach, we do not make any distributional assumption on the birth weights. We define the percentile-specific treatment effect Δ_p^W as the difference between the quantile functions of the birth weights for the treatment and the control, and we assume that such difference is a smooth function of

the percentiles of the birth weight distribution. That is:

$$\Delta_p^W = Q_1(p) - Q_0(p) = s(p, \lambda) \tag{2}$$

where s is a natural cubic spline of the percentile p with λ degrees of freedom (we set $\lambda = 5$).

To estimate Δ_p^W for $0 < p < 1$, we:

1. calculate the percentiles $p_i = i/(n_0 + 1)$ with $n_0 = 766$ (the smallest number of infants across treatment groups);
2. calculate the differences between the empirical quantiles of the birth weights $\widehat{Q}_1(p_i) - \widehat{Q}_0(p_i)$;
3. smooth these differences across the percentiles p_i.

Note that for $p = 0.5$, estimating $\Delta_{p=0.5}^W$ reduces to the usual method of estimating a treatment effect by comparing medians between the treatment and control groups.

To account for the uncertainty in the imputation of the missing values, we repeated steps 1-3 separately for 50 imputed data sets. We then calculate the percentile-specific treatment effect and its corresponding total statistical variance by using standard multiple imputation methods (Rubin 1987). Let $\widehat{\Delta}_p^{W^{(j)}}$ and $V^{(j)}(p)$ be the point estimate and the bootstrap variance of Δ_p^W for the j-th imputed data set, respectively. For each j, we obtain the overall estimate of the treatment effect and its total variance, denoted by $\widehat{\Delta}_p^W$ and $\widehat{TV}_p$, as follows:

$$\begin{array}{lcl} \widehat{\Delta}_p^W & = & \frac{1}{J}\sum_{j=1}^J \widehat{\Delta}_p^{W^{(j)}} \\ \widehat{TV}_p & = & A_p + (1 + \frac{1}{J})B_p, \text{ where} \\ A_p & = & \frac{1}{J}\sum_{j=1}^J V_p^{(j)} \\ B_p & = & \frac{1}{J-1}\sum_{j=1}^J (\widehat{\Delta}_p^{W^{(j)}} - \widehat{\Delta}_p^W)^2. \end{array} \tag{3}$$

Permutation test: Finally, to test whether the treatment effect is constant across the percentiles of the birth weight distribution, we perform a permutation test. Specifically, for $h = 1, \ldots, 500$, we randomly re-assign the birth weights to the two treatment groups and calculate the test statistics $T^h = \sum_{i=1}^{n_0} (\widehat{s}^h(p_i, \lambda) - \bar{s}^h)^2$ where $\bar{s}^h = \sum_{i=1}^{n_0} \widehat{s}^h(p_i, \lambda)$. We calculate the one-sided p-value as the probability that T^h exceeds the observed test statistics $T_{obs} = \sum_{i=1}^{n_0} (\widehat{s}(p_i, \lambda) - \bar{s})^2$ where $\bar{s} = \sum_{i=1}^{n} \widehat{s}(p_i, \lambda)$.

The modelling approach illustrated in this section has been described elsewhere (Katz et al. 2005). The idea of smoothing quantile differences across percentiles to improve estimation of the average difference between two outcomes has recently been discussed by Dominici et al. (2005a) for estimating the difference in means for skewed distributions. This approach was then implemented for estimating average medical expenditures between diseased and non-diseased patients (Dominici and Zeger 2005).

In this paper we have tailored this idea for the ultimate goal of estimating percentile-specific treatment effects.

4 A Bayesian Model with Measurement Error

In this section, we define a Bayesian approach for approximating the marginal posterior distributions of all parameters of scientific interest accounting for 1) measurement error in the birth weights; 2) uncertainty in the imputation of the missing values; and 3) uncertainty in the imputation of the missing counterfactuals.

Adopting a counterfactual model (Rubin 1978; Holland 1986), let Z_i be the treatment assignment, and $W_i(Z_i)$ be the birth weight of baby i given the treatment assignment Z_i. We define $Y_i(Z_i)$ to be the mortality indicator for baby i corresponding to treatment assignment Z_i. We refer to $Y_i(Z_i)$ and $W_i(Z_i)$ as potential outcomes. Note that $Y_i(0)$ and $W_i(0)$ are defined for all N babies, but only observed for the n_0 babies in the control group of the study. Similarly, $Y_i(1)$ and $W_i(1)$ are defined for all N babies, but only observed for the n_1 babies in the intervention group. Thus $Y_i^{obs} = \{Y_i(z),\ \text{if } z = Z_i\}$ and $W_i^{obs} = \{W_i(z),\ \text{if } z = Z_i\}$, respectively. Finally, let t_i be the time at which birth weight is measured for baby i. Since weights are stable in the first 72 hours, we define $t_i = 0$ for the interval 0-72. Let $W_{it_i}(Z_i)$ be the potential weight at time t_i.

We define the likelihood function for the complete data as a function of three vectors of unknown parameters:

$$\begin{aligned} L(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\eta}_3) &= \prod_{i=1}^{N} Pr(Y_i(1), Y_i(0) \mid W_i(1), W_i(0), \boldsymbol{\eta}_1) \times f_1(W_i(1), W_i(0) \mid \boldsymbol{x}_i, \boldsymbol{\eta}_2) \times \\ &\times \prod_{i \in M} f_2(W_{it_i}(0), W_{it_i}(1) \mid W_i(0), W_i(1), t_i, \boldsymbol{\eta}_3). \end{aligned} \tag{4}$$

In the next three subsections we specify: 1) an odds-ratio association model for bivariate mortality indicators given the birth weights $P(Y_i(1), Y_i(0) \mid W_i(1), W_i(0), \boldsymbol{\eta}_1)$ (Liang et al. 1992); 2) the joint distribution of $f_1(W_i(1), W_i(0) \mid \boldsymbol{x}_i, \boldsymbol{\eta}_2)$ as a bivariate normal given the mother's covariates; and 3) the measurement error model for the babies weighted after 72 hours $f_2(W_{it_i}^{obs} \mid W_i^{obs}, t_i, \boldsymbol{\eta}_3)$.

4.1 Statistical model for infant mortality given birth weight

Following Liang et al. (1992), we parameterize the 2×2 joint distribution $[Y_i(0), Y_i(1) \mid W_i(0), W_i(1), \boldsymbol{\eta}_1]$ in terms of the two margins and the odds ratio. Specifically, we assume that:

$$\begin{aligned} P\{Y_i(0) = y_i(0), Y_i(1) = y_i(1) \mid W_i(0), W_i(1), \boldsymbol{\eta}_1\} &= \\ \mu_i(0)^{y_i(0)}(1-\mu_i(0))^{1-y_i(0)} &\times \mu_i(1)^{y_i(1)}(1-\mu_i(1))^{1-y_i(1)} + \\ (-1)^{y_i(0)-y_i(1)}\{\mu_i(11) - \mu_i(0)\mu_i(1)\} & \end{aligned} \tag{5}$$

where $\mu_i(1) = Pr(Y_i(Z_i) = 1 \mid Z_i, W_i(Z_i))$ is assumed to follow the logistic model:

$$\text{logit}Pr\{Y_i(Z_i) = 1 \mid Z_i, W_i(Z_i)\} = \beta_0 + \beta_1 Z_i + s(W_i(Z_i), 3),\ Z_i = 0, 1, \qquad (6)$$

and $s()$ denotes a natural cubic spline with 3 knots. The parameter $\mu_i(11) = Pr(Y_i(0) = Y_i(1) = 1 \mid W_i(0), W_i(1))$ is a known function of the marginal probabilities $\mu_i(1), \mu_i(0)$ and of the pre-specified odds ratio ψ. Thus $\boldsymbol{\eta}_1 = (\boldsymbol{\beta}, \psi)$ where $\boldsymbol{\beta}$ also includes the regression coefficients of the spline basis.

Let $I_0 \subset I$ and $I_1 \subset I$ be the two sub-populations of n_0 and n_1 infants assigned to the control and to the treatment, respectively. Within Gibbs sampling we will sample the missing counterfactuals from the conditional distributions $[Y_i(0) \mid Y_i(1), W_i(1), W_i(0), \boldsymbol{\eta}_1]$ for $i = 1, \ldots, n_1$ and from $[Y_i(1) \mid Y_i(0), W_i(1), W_i(0), \boldsymbol{\eta}_1]$ for $i = 1, \ldots, n_0$. Note that this imputation depends upon unverifiable assumptions about the association between the counterfactual pairs of variables $\{Y_i(0), Y_i(1)\}$ denoted by the parameter ψ. We assume that ψ is known and we will perform sensitivity analyses with respect to different values for ψ. The rationale behind the range of values considered is provided in section 4.4.

4.2 Statistical model for birth weight

We specify the joint distribution $f_1(W_i(1), W_i(0) \mid \boldsymbol{x}_i, \boldsymbol{\eta}_2)$ as follows:

$$\begin{pmatrix} W_i(0) \\ W_i(1) \end{pmatrix} \sim N_2 \left(\begin{pmatrix} \alpha_{00} + \boldsymbol{\alpha}_0(\boldsymbol{x}_i - \bar{\boldsymbol{x}}) \\ \alpha_{01} + \boldsymbol{\alpha}_1(\boldsymbol{x}_i - \bar{\boldsymbol{x}}) \end{pmatrix}, \begin{bmatrix} \sigma_0^2 & \sigma_0\sigma_1\rho \\ \sigma_0\sigma_1\rho & \sigma_1^2 \end{bmatrix} \right),\ i = 1, \ldots, N \qquad (7)$$

where

$$\begin{aligned} \alpha_{0z} + \boldsymbol{\alpha}_z(\boldsymbol{x}_i - \bar{\boldsymbol{x}}) \quad = \quad & \alpha_{0z} + \alpha_{1z}\text{num.cig}_i + \alpha_{2z}\text{gest.age}_i + \alpha_{3z}\text{mom.weight}_i + \\ & + \alpha_{4z}\text{mom.height}_i + \alpha_{5z}\text{mom.age}_i,\ z = 0, 1. \end{aligned} \qquad (8)$$

Thus $\boldsymbol{\eta}_2 = (\alpha_{0z}, \boldsymbol{\alpha}_z, z = 0, 1, \sigma_0, \sigma_1, \rho)$.

Under model (7) and within the Gibbs sampling, we will carry out two types of imputation. The first imputation borrows strength across babies and use the mother's covariates to impute the missing birth weights. Let n_{0mis} and n_{1mis} be the number of missing birth weight measurements for the control and treated groups where $I_{mis} = I_{0mis} \bigcup I_{1mis}$ and $n_{mis} = n_{0mis} + n_{1mis}$. At each iteration of the Gibbs sampling, we will sample: 1) the missing birth weights for the control group from the full conditional distribution $[W_i(0) \mid Y_i(0), \boldsymbol{x}_i, \boldsymbol{\eta}_2]$ for $i \in I_{0mis}$ and 2) the missing birth weights for the treatment group from the full conditional distribution $[W_i(1) \mid Y_i(1), \boldsymbol{x}_i, \boldsymbol{\eta}_2]$ for $i \in I_{1mis}$.

The second imputation relies on the correlation ρ between $W_i(0)$ and $W_i(1)$ for the same baby to impute the missing counterfactuals. That is, we will impute the missing counterfactuals by sampling from the full conditional distribution $[W_i(0) \mid W_i(1), Y_i(0), Y_i(1), \boldsymbol{\eta}_2]$ for $i \in I_1$ and from $[W_i(1) \mid W_i(0), Y_i(0), Y_i(1), \boldsymbol{\eta}_2]$ for $i \in I_0$. Note that this second imputation depends upon unverifiable assumptions about ρ. Like

for ψ, we assume that ρ is known but we perform sensitivity analyses of our results with respect to different values for ρ.

4.3 Measurement Error Model

In this section we specify a measurement error model that allows us to sample the "birth weights" for the infants that have been measured after 72 hours. Let M_0 and M_1 be the subsets of m_0 and m_1 infants that have been measured after 72 hours under the control and the treatment groups respectively. We assume that:

$$\begin{aligned}\prod_{i\in M} f_2(W_{it_i}(0), W_{it_i}(1) \mid W_i(0), W_i(1), t_i, \boldsymbol{\eta}_3) &= \prod_{i\in M_0} f_2(W_{it_i}(0) \mid W_i(0), t_i, \boldsymbol{\eta}_3) \times \\ &\quad \prod_{i\in M_1} f_2(W_{it_i}(1) \mid W_i(1), t_i, \boldsymbol{\eta}_3).\end{aligned}$$

That is, we assume that:

1. the measurements made after 72 hours are independent across treatment groups conditionally on the birth weights:

$$\begin{aligned}[W_{it_i}(0), W_{it_i}(1) \mid W_i(0), W_i(1), t_i, \boldsymbol{\eta}_3] &= [W_{it_i}(0) \mid W_i(0), W_i(1), t_i, \boldsymbol{\eta}_3] \times \\ &\quad [W_{it_i}(1) \mid W_i(0), W_i(1), t_i, \boldsymbol{\eta}_3];\end{aligned}$$

2. the measurements made after 72 hours depend only on the birth weights for the same treatment group, that is:

$$[W_{it_i}(Z_i) \mid W_i(Z_i), W_i(1-Z_i), t_i, \boldsymbol{\eta}_3] = [W_{it_i}(Z_i) \mid W_i(Z_i), t_i, \boldsymbol{\eta}_3].$$

We then specify the following measurement error model:

$$W_{it_i}(z) \mid t_i \sim N\left(\gamma_0 + \gamma_1 t_i, \tau^2\right), i \in I \backslash I_{mis}. \tag{9}$$

Ideally, we would like to allow each baby to have his/her own random intercept. However, because we have only one birth weight measurement for each baby, a random intercept model is not identifiable.

Let $\gamma_0^{(j)}$ be the current value of γ_0 at the j-th iteration of the Gibbs sampler. We then sample the weight at time 0 for the infants with $t_i > 72$ as following:

$$\begin{aligned}W_i^{(j)}(0) &= \gamma_0^{(j)} + \delta_i(0),\ i \in M_0 \\ W_i^{(j)}(1) &= \gamma_0^{(j)} + \delta_i(1),\ i \in M_1\end{aligned}$$

where $\delta_i(z_i)$ is a known quantity. We obtain the value of $\delta_i(z_i)$ outside the Gibbs sampling by fitting a regression on $(W_{it_i}(z_i), t_i)$ for $i \in M$ and setting $\delta_i(z_i)$ equal to $\widehat{W}_{it_i}(z_i) - W_{it_i}(z_i)$ where $\widehat{W}_{it_i}(z_i)$ are the predicted values.

4.4 Parameters of Scientific Interest

Some parameters of interest are defined in Table 3. The first row of Table 3 defines the average counterfactual treatment effect on birth weight. The second row defines the percentile-specific treatment effects on birth weight. Note that the parameter Δ_p^W is defined as a function of the marginal distributions of $W_i(1)$ and $W_i(0)$ and therefore it does not depend on the parameter ρ. In addition, the distributional assumption (7) allows the parameter Δ_p^W to vary flexibly but smoothly as a function of the percentiles (p) of the birth weight distribution.

If we do not account for the mother's covariate and we assume

$$\begin{pmatrix} W_i(0) \\ W_i(1) \end{pmatrix} \sim N_2 \left(\begin{pmatrix} \mu_0 \\ \mu_1 \end{pmatrix}, \begin{bmatrix} s_0^2 & s_0 s_1 \rho \\ s_0 s_1 \rho & s_1^2 \end{bmatrix} \right), \; i = 1, \ldots, N \tag{10}$$

then $\Delta_p^W = Q_1(p) - Q_0(p) = (\mu_1 - \mu_0) + \Phi^{-1}(s_1 - s_0)$, and if we further assume that $s_1 = s_0$, then Δ_p^W is not allowed to vary with p.

Throughout the paper we will compare our posterior inferences on Δ_p^W under three models. The first model, denoted as Model A and also defined in Equation (7), accounts for the mother's covariate and the uncertainty in the imputation of the missing birth weights. The second model, denoted as Model B and defined in Equation (10), uses the "working data set" and ignores uncertainty in the imputation of missing birth weights and prediction of birth weights measured after the 72 hours. The third model, denoted as Model C, is the non-parametric model with multiple imputation discussed in Section 3. In addition, we will estimate the tail probabilities of the distribution $\log(s_1^2/s_0^2)$ under (10) to provide evidence to assess whether the treatment effect varies as a function of the birth weight percentiles. We will compare these posterior probabilities with the p-values obtained from the permutation test described in Section 3.

The rest of Table 3 summarizes the parameters of scientific interest for the treatment effects on infant mortality. The third row indicates the average "counterfactual" treatment effect on survival. The fourth row introduces the percentile-specific effects of treatment on survival defined as the difference in the probability of death between treated and non-treated infants who are at the same percentiles of their respective birth weight distribution. Note that this parameter is defined as a function of the marginal distributions of $Y_i(0) \mid W_i(0)$ and $Y_i(1) \mid W_i(1)$ and therefore does not depend on ψ.

In the last four rows of Table 3, we implement the idea of principal stratification by Frangakis and Rubin (2002) for defining causal parameters of the effects of treatment on infant mortality that are "adjusted" and "mediated" by post-treatment changes in birth weight. More specifically, τ_1^Y and τ_2^Y are the effects of treatment on mortality in the two sub-populations of LBW babies for whom the treatment effect on birth weight was smaller and larger than 50 grams, respectively. Thus a comparison between τ_1^Y and τ_2^Y measures the degree to which a causal effect of treatment on mortality occurs together with a causal effect of treatment on the birth weight among the LBW infants. The parameters τ_3^Y and τ_4^Y are the analogues of τ_1^Y and τ_2^Y for the not-LBW infants,

that is, for the infants with birth weight larger than 2500 grams.

The average effects obtained under the counterfactual model may depend upon unverifiable assumptions about the joint distribution of the counterfactual pairs of variables $\{W_i(0)$ and $W_i(1)\}$, and $\{Y_i(0)$ and $Y_i(1)\}$. As anticipated in the previous section, in order to estimate these parameters, we make the following key but unverifiable assumptions about the correlation between the observed outcomes and their counterfactuals. First, we assume that the correlation between $W_i(Z_i)$ and $W_i(1-Z_i)$, denoted by ρ, is known and equal to 0.9. We will perform sensitivity analyses for $\rho = 0.5$. Second, we assume that the odds ratio between the observed and counterfactual mortality given birth weight, denoted by ψ,is equal to 25. We will perform sensitivity analyses for $\psi = 1.5$. These choices have been guided by exploratory analyses of data from this randomized trial and from other data sources (Rahmathullah et al. 2003; Katz et al. 2000b, 2001) which have been used to estimate the correlations of birth weights for two successive children born to the same mother and birth weights for twins.

5 Computation

To investigate the posterior distributions of all parameters of interest we implement a Monte Carlo Markov Chain method with data augmentation for imputing the missing data (Tanner 1991; Gelman et al. 1995). We implemented a Metropolis-within-Gibbs (Tierney 1994) approach, in which both the parameters and the counterfactual variables are sampled using a random walk proposal. Computational details and full conditionals are summarized in the Appendix. We specify flat prior distributions on all the unknown parameters, except for the parameters ρ and ψ which are equal to pre-specified fixed values.

For each posterior sample of the unknown parameters and counterfactuals, we obtain a posterior sample of the percentile-specific parameters as follows. To obtain a posterior sample of Δ_p^W, we sort $W_i(0)$ and $W_{i'}(1)$ within the two groups of treated and untreated babies separately, and then we take their difference. Under model (10) we obtain a posterior sample of Δ_p^W by using the posterior samples of the parameters of the joint normal distributions and plotting the theoretical function $\mu_1 - \mu_0 + \Phi^{-1}(p)(s_1 - s_0)$.

To calculate a posterior sample of Δ_p^Y, we first sort sample values of $Y_i(0)$ with respect to $W_i(0)$ and $Y_{i'}(1)$ with respect to $W_{i'}(1)$ within each of the two groups separately, and then we take the difference. We smoothed the posterior samples of these percentile-specific parameters to reduce Monte Carlo variability in the posterior probability bounds.

6 Results

Figure 2 shows birth weights plotted versus times of measurement. Red dots denote birth weights measured under the treatment and green dots denote birth weights measured under the control. The segments connect a random subset of the observed measurements

$W_{it_i}^{obs}$ to the Bayesian posterior means of the predicted measurements at time zero W_i^{obs} for $i \in M$.

Figure 3 shows the marginal posterior distributions of the average treatment effect $TE^W = E[W_i(1) - W_i(0)]$ under two model specifications: 1) Model A defined in Equation (7): a Bayesian model that accounts for the uncertainty in the imputation of the missing data, the estimation of the birth weights at time zero, and the mother's covariates (red curve); 2) Model B defined in Equation (10): a Bayesian model that uses one imputed data set only and that does not account for the mother's covariates (green curve). Overall we found that both supplementations are effective and increase birth weight. Under Model A we obtain a smaller estimate of the average causal treatment effect than under Model B. As expected, posterior inferences under Model A lead to an estimate with larger posterior intervals than Model B because Model A accounts for the uncertainty in the imputation of the missing birth weights and in the prediction of the measurements after 72 hours.

Figure 4 shows the marginal posterior distributions of the percentile-specific treatment effects on birth weight (Δ_p^W) under Models A and B (red and green curves) described above and under Model C (blue curve), a non-parametric model for the birth weights with multiple imputation for the missing data (see Section 2). The grey polygon denotes the corresponding 95% posterior confidence bands under Model A. The green curve is obtained by taking the point-wise posterior means of the theoretical function $\Delta_p^W = \mu_1 - \mu_0 + \Phi^{-1}(s_1 - s_0)$. At the far right are shown the point estimates and 95% uncertainty bands of the average treatment effect $E[W_i(1)] - E[W_i(0)]$ under the three models. Inferences are similar across models.

In previous work (Dominici et al. 2005b), we have also modeled the joint distribution of the birth weights in a more flexible way, by assuming that the margins follow a mixture of three normal distributions and by introducing a correlation parameter ρ between the standardized variables $\Phi^{-1}[F_0(W_i(0))]$ and $\Phi^{-1}[F_1(W_i(1))]$, where Φ is the cdf of a standard normal distribution and F_0, F_1 are the cdf of a mixture of three normal distributions of $W_i(0)$ and $W_i(1)$ respectively. We found that results under this mixture model were very similar to the simpler ones shown here.

Although the two micronutrient supplementations have similar average causal effects, their percentile-specific treatment effects differ substantially. In Panel (a), for the F+I+A group, the estimated Δ_p^W are decreasing functions of p indicating that the estimated treatment effects decrease from more than 100 grams in the left tail to 0 grams in the right tail. In Panel (b), for the M+A group, these parameters are almost a constant function of p. Under Model B, the posterior probability that $\log s_1^2 - \log s_0^2$ is less than zero is 97% in Panel (a) and 70% in Panel (b). We have strong evidence of an interaction between the treatment effect and the percentiles of the birth weight distribution for F+I+A but not for M+A. Under Model C, we found that the one-sided p-values from the permutation test described in Section 2 were equal to 0.10 for F+I+A and equal to 0.96 for M+A.

Figure 5 shows the posterior means and 95% posterior regions of the percentile-specific difference in infant mortality rates between the treatment and control popula-

tions (Δ_p^Y) plotted with respect to the percentiles of the birth weight distributions. For a specific p, Δ_p^Y is the difference in the probability of death between two babies with birth weights $W_i(1), W_{i'}(0)$, each at the p-percentile of their respective birth weight distributions. The vertical dotted line is placed at the 0.42 percentiles corresponding to 2500 grams in the control sample. For F+I+A, there is suggestive evidence that the treatment reduces mortality among the smallest babies but has no benefit for the babies above the median birth weight. For M+A, these posterior inferences suggest that the treatment does not affect mortality and that it might actually slightly increase the risk among the largest babies.

Figure 6 shows posterior distributions of the average treatment effects on mortality separately for five sub-populations of infants. These boxplots also show the sensitivity of our posterior inferences to specification of the values for the parameters ρ and ψ. The first set of boxplots (posterior distributions of τ_1^Y) indicate that, among the LBW babies with little change in birth weight after the supplementation, there is no evidence that F+I+A supplementations affect survival and weak evidence that the M+A might increase the risk of death. For the F+I+A (Panel a), the second set of four boxplots (posterior distributions of τ_2^Y) suggest that, among the LBW babies with increase in birth weight after the supplementation larger than 50 grams, there is strong evidence that this intervention is beneficial. For M+A (Panel b), this evidence is less strong. The third set of boxplots (posterior distributions of the τ_3^Y) indicate that, among the no-LBW babies with little change in birth weight after the supplementation, we found no evidence that the F+I+A supplementation is associated with survival. However, for M+A (Panel b) we found evidence that this intervention might increase the risk of death. The fourth set of boxplots (posterior distributions of τ_4^Y) suggests no evidence of an association between both supplementations and survival. Finally, overall for the entire population of babies (last set of boxplots), we found weak evidence that F+I+A improves survival. Results are not sensitive to the choice of ρ and ψ.

We then estimated the posterior distribution of the proportion of infants (p) whose survival is improved by the F+I+A supplementation. Specifically, we define the m-th posterior sample of p as

$$\widehat{p}^{(m)} = \frac{1}{N}\sum_{i=1}^{N} \#\left(W_i^{(m)}(1) - W_i^{(m)}(0) > 50 \;\&\; W_i^{(m)}(0) \leq 2500\right),$$

where i is the child. We found that the posterior mean $\frac{1}{M}\sum_{m=1}^{M}\widehat{p}^{(m)} = 0.27$ indicating that 27% of the infants are in the stratum $W_i(1) - W_i(0) > 50$ & $W_i(0) \leq 2500$.

Finally we investigate whether the predictions of principal strata are sharp enough that, in the future, supplements might be given only to mothers whose infants are predicted to be in the $W(0) \leq 2500$, $W(1) - W(0) > 50$ strata. Specifically, for each child we estimate the vector of probabilities $\boldsymbol{\pi}_i = (\pi_{i1}, \pi_{i2}, \pi_{i3}, \pi_{i4})$, where π_{ij} is the probability that infant i is in the principal stratum $j = 1, 2, 3, 4$. That is, we estimate

the π_{i2} as:

$$\widehat{\pi}_{i2} = \frac{1}{M}\sum_{m=1}^{M} \#\left(W_i^{(m)}(1) - W_i^{(m)}(0) > 50 \;\&\; W_i^{(m)}(0) \leq 2500\right).$$

Figure 7 shows histograms of the probabilities $\widehat{\pi}_{i1} + \widehat{\pi}_{i2} = P(W_i(0) \leq 2500)$ (Panel a), $\widehat{\pi}_{i2}/(\widehat{\pi}_{i1} + \widehat{\pi}_{i2}) = P(W_i(1) - W_i(0) > 50 \mid W_i(0) \leq 2500)$ (Panel b), and $\widehat{\pi}_{i4}/(\widehat{\pi}_{i3} + \widehat{\pi}_{i4}) = P(W_i(1) - W_i(0) > 50 \mid W_i(0) > 2500)$ (Panel c), respectively. The vertical lines are placed at 0.2 and 0.8.

Predictions of children in the principal strata are reasonable sharp. In Panel (a) we found that 45% and 33% of the children have $\widehat{\pi}_{i1} + \widehat{\pi}_{i2}$ less than 0.2 and larger than 0.8. In Panel (b) we found that 0% and 40% of the children have $\widehat{\pi}_{i2}/(\widehat{\pi}_{i1} + \widehat{\pi}_{i2})$ less than 0.2 and larger than 0.8. Finally in Panel (c) we found that 26% and 0.003% have $\widehat{\pi}_{i4}/(\widehat{\pi}_{i3} + \widehat{\pi}_{i4})$ less than 0.2 and larger than 0.8.

In summary, these results indicate that F+I+A has an effect, where it most need, of increasing the birth weight among the LBW infants and increasing their chances of survival. Instead the M+A intervention, because it increases the birth weight among the not-LBW, is a less ideal intervention than the F+I+A and might harm the largest babies.

7 Discussion

A micronutrient supplementation trial is considered effective if the treatment reduces the risk of infant mortality either directly or through increases in birth weight. Because infant mortality is greatest among low birth weight infants (LBW), an intervention to increase fetal growth must increase birth weight mainly among the smallest babies. A community-based trial in Nepal has shown that a multiple micronutrient supplementation increases birth weight but the limitation in the study size has to date prevented us from unambiguously establishing that this translates into a mortality benefit (Christian et al. 2003b).

Our analysis demonstrates that the standard approach of estimating a mean difference in a continuous outcome between a treatment and control group may not adequately capture the impact of nutritional supplementation on birth weight. The ability to assess whether the treatment effect varies across the distribution of the outcome may provide insights into the mechanism by which the treatment affects the outcome, and ideas as to why a surrogate outcome (such as birth weight) may not reflect the effect of treatment on the real outcome of interest (mortality).

In this paper, we develop a counterfactual model to evaluate the efficacy of micronutrient supplementation trials in developing countries. We focus on whether the supplementation increases birth weight and ultimately survival differently among the smaller and the larger babies, and whether the supplementation improves survival largely through its positive effect on birth weight or it improves survival even without

affecting the birth weight. This analysis demonstrates that inference about counterfactual treatment effects in the middle of the birth weight distribution are relatively robust to unverifiable assumptions about the joint distribution of the counterfactuals. However, in our previous work (Dominici et al. 2005b), we have provided evidence that inference about counterfactual treatment effects on birth weights at the tails of the birth weight distribution are sensitive to these unverifiable assumptions.

The posterior distributions of all the parameters are evaluated by using Bayesian inferences with data-augmentation methods
(Tanner and Wong 1987; Tanner 1991; Albert and Chib 1993; Chib and Greenberg 1998).
A nice feature of this approach is that we can evaluate the posterior distributions of the quantities of interest taking into account uncertainty in the imputation of the the missing counterfactuals, missing data and measurement error. In addition, we can explore the sensitivity of the posterior inferences to unverifiable assumptions about the joint distribution between the observed and the counterfactual variables.

For estimating percentile-specific effects of the treatment on birth weights, we developed and compared three modelling approaches for the difference in quantile functions: 1) model A assumes that $(W_i(0), W_i(1))$ is jointly normal with marginal means that depend on the mother's covariate profile, and we fit this model accounting for the uncertainty in the imputation of the missing birth weights and in the prediction of the birth weights for the infants that were measured after 72 hours; 2) Model B assumes that $W_i(0), W_i(1)$ is jointly normal but with marginal means (μ_0, μ_1) that do not depend on the mother's covariates, and we fit this model by relying on one "working" data set where the missing data and the measurements made after 72 hours where replaced by predicted values from a regression model (9); and 3) Model C which simply assumes that the quantile function difference is a smooth function of the percentiles. Missing data were imputed by use of multiple imputation. These three models provided very similar results on the average treatment effects.

In summary, we have provided an inferential framework for estimating treatment effects in counterfactual models in a randomized trial with a continuous post-treatment variable. By comparing population with counterfactual parameter estimates, carrying out sensitivity analyses, and implementing principal stratification, we have characterized the amount of evidence supporting the scientific questions of interest and their sources of uncertainty.

We found that the treatment effects varied across the birth weight distribution for F+I+A but not for M+A. In fact, there was a constant treatment effect of the M+A of about 90 grams. For F+I+A, the average treatment effect was 100 grams at the lower end of the distribution. In environments like rural Nepal, it may be more important to selectively affect the lower rather than the upper part of the birth weight distribution. In fact, impacting the upper part of the distribution may be harmful to the mother and infant.

We found the multiple micronutrient supplement to be associated with a slightly elevated risk of early infant mortality, especially among the no-LBW infants, although with large statistical uncertainty. This was despite the significant increase in birth

weight. The risk of birth asphyxia as a cause of neonatal mortality also appeared to be higher in the group receiving the multiple micronutrient supplement. On the other hand, folic acid plus iron was associated with an overall reduction of infant mortality among LBW-infants. Given an improvement in birth weight at the lower end of the distribution, this intervention may have produced improved survival overall, while the multiple micronutrient appeared to have no impact on survival because deaths averted in the smaller infants were negated by higher mortality at the upper end of the distribution.

The estimation of treatment effects by percentile of the birth weight distribution has public health significance. We have carried out additional analyses to investigate whether predictions of infants in principal strata are reasonably sharp. For the F+I+A we found that 33% of the infant have a probability of being low birth weight larger than 0.80 and that 40% of the low birth weight infants have a probability $W_i(1) - W_i(0) > 50$ larger than 0.80. Therefore from a public health perspective, this approach can also help identify whether a targeted, rather than universal supplementation program would be more effective and efficient in achieving a nutritional goal for a population. We can use covariate information to predict those mothers who are likely to have larger infants and to exclude them from intervention programs.

Currently, recommendations exist for supplementing women with iron-folic acid during pregnancy in developing countries. The Nepal study (Christian et al. 2003a) demonstrates that beyond reducing anemia, iron can result in an improvement in birth weight primarily through moving the lower tail of the birth weight distribution to the right. Presumably, this effect is mediated through improving the iron status of those pregnant women who are the most iron deficient. These data from Nepal reveal that when evaluating public health interventions it is important to be, at the very least, cognizant of the different beneficial effects of an intervention depending on where in the distribution the program participants fall and that an overall effect size may: 1) under-estimate the maximum likely benefit in the most malnourished individuals; and 2) incorrectly assume benefits where none exist and potentially mask harm in the more well-nourished individuals.

Appendix

List of full conditionals in the Gibbs sampling

- missing birth weights: $[W_i(0) \mid Y_i(0), \boldsymbol{x}_i, \boldsymbol{\eta}_2]$ for $i \in I_{0mis}$ and $[W_i(1) \mid Y_i(1), \boldsymbol{x}_i, \boldsymbol{\eta}_2]$ for $i \in I_{1mis}$. These are not available in closed form and we implement a metropolis step;

- birth weights for the measurements made after the 72 hours: $[W_i(1) \mid W_{it_i}(1), t_i, Y_i(1), \boldsymbol{\eta}_3]$ for $i \in M_1$ and from $[W_i(0) \mid W_{it_i}(0), t_i, Y_i(0), \boldsymbol{\eta}_3]$ for $i \in M_0$ respectively. These are not available in closed form and we implement a metropolis step;

- missing counterfactuals for the birth weights: $[W_i(0) \mid W_i(1), Y_i(0), Y_i(1), \boldsymbol{\eta}_2]$ for

$i \in I_1$ and from $[W_i(1) \mid W_i(0), Y_i(0), Y_i(1), \boldsymbol{\eta}_2]$ for $i \in I_0$. These are not available in closed form and we implement a metropolis step;

- missing counterfactuals for the mortality indicators: $[Y_i(0) \mid Y_i(1), W_i(1), W_i(0), \boldsymbol{\eta}_1]$ for $i \in I_1$ and from $[Y_i(1) \mid Y_i(0), W_i(1), W_i(0), \boldsymbol{\eta}_1]$ for $i \in I_0$. These are not available in closed form and we implement a metropolis step;

- we generate γ_0 from the full conditional distribution:

$$N\left(\frac{1}{N} \times (\sum_i t_i(W_{it_i}(Z_i) - \gamma_1 t_1); \frac{1}{N} \times \tau^2\right);$$

- we generate γ_1 from the full conditional distribution:

$$N\left(\frac{1}{\sum_i t_i^2} \times (\sum_i t_i(W_{it_i}(Z_i) - \gamma_0); \frac{1}{\sum_i t_i^2} \times \tau^2\right);$$

- we generate τ^2 from the full conditional distribution:

$$IG\left(N/2 - 1; \frac{1}{2}\sum_i (W_{it_i}(Z_i) - \gamma_0 - \gamma_1 t_i)^2\right);$$

- we generate $\boldsymbol{\alpha}_0$ from the full conditional

$$N_p\left([\sum_i \boldsymbol{x}_i^{'}\boldsymbol{x}_i]^{-1} \times \sum_i \boldsymbol{x}_i^{'} W_i(0); V_0\right), \text{ where } V_0 = \left[\frac{1}{\sigma_0^2}\sum_i \boldsymbol{x}_i^{'}\boldsymbol{x}_i\right]^{-1};$$

- we generate $\boldsymbol{\alpha}_1$ from the full conditional

$$N_p\left([\sum_i \boldsymbol{x}_i^{'}\boldsymbol{x}_i]^{-1} \times \sum_i \boldsymbol{x}_i^{'} W_i(1); V_1\right), \text{ where } V_1 = \left[\frac{1}{\sigma_1^2}\sum_i \boldsymbol{x}_i^{'}\boldsymbol{x}_i\right]^{-1};$$

- the full conditionals of σ_0^2 and σ_1^2 are not available in closed form. We implement a metropolis step where the proposal distribution is log-normal with mean equal to the logarithm of the current value of the parameter and known variance;

- the full conditional of $\boldsymbol{\beta}$ is not available in closed form. We implement a metropolis step where the proposal distribution is multivariate normal with mean equal to the current value of the parameter and covariance matrix obtained by fitting the logistic regression model (6) to the data.

References

Albert, J. H. and Chib, S. (1993). "Bayesian Analysis of Binary and Polychotomous Response Data." *Journal of the American Statistical Association*, 88: 669–679. 2, 5, 17

Caulfield, L., Zavaleta, N., Figueroa, A., and Leon, Z. (1999). "Maternal Zinc Supplementation does not affect size at birth and pregnancy duration in Peru." *Journal of Nutrion*, 129: 1563–8. 3

Ceesay, S., Prenctice, A., Cole, T., Foord, F., Weaver, L., and Poskitt, E. e. a. (1997). "Effects on birth weight and perinatal mortality of maternal dietary supplements in rural Gambia: 5 randomized controlled trials." *British Medical Journal*, 315: 786–790. 3

Chib, S. and Greenberg, E. (1998). "Analysis of Multivariate Probit Models." *Biometrika*, 85: 347–361. 2, 5, 17

Christian, P., Khatry, S., Katz, J., Pradhan, E., LeClerq, S., Shrestha, S., Adhikari, R., Sommer, A., and West, K. (2003a). "Effects of alternative maternal micronutrient supplements on low birth weight in rural Nepal: double blind randomised community trial." *British Medical Journal*, 326: 1–6. 3, 18

Christian, P., West, K., Khatry, S., Leclerq, S., Pradhan, E., Katz, J., Shrestha, S., and Sommer, A. (2003b). "Effects of maternal micronutrient supplementation on fetal loss and infant mortality: a cluster-randomized trial in Nepal." *American Journal of Clinical Nutrition*, 78: 1194–1202. 3, 5, 16

Dominici, F., Cope, L., Naiman, D., and Zeger, S. L. (2005a). "Smooth Quantile Ratio Estimation (SQUARE)." *Biometrika (to appear)*, 92: ?–? 8

Dominici, F. and Zeger, S. (2005). "Smooth Quantile Ratio Estimation with Regression: Estimating Medical Expenditures for Smoking Attributable Diseases." *Biostatistics (to appear)*. 8

Dominici, F., Zeger, S., Parmigiani, G., Katz, J., and Christian, P. (2005b). "Estimating Percentile-Specific Causal Effects: A Case Study of Micronutrient Supplementation, Birth Weight, and Infant Mortality." Technical report, Department of Biostatistics Johns Hopkins University. 5, 14, 17

Frangakis, C. E. and Rubin, D. B. (2002). "Principal Stratification in Causal Inference." *Biometrics*, 58(1): 21–29. 2, 5, 12

Garner, P., Kramer, M., and Chalmers, L. (1992). "Might efforts to increase birth weight in undernourished women do more harm than good?" *Lancet*, 340: 1021–1022. 3

Gelman, A., Carlin, J., Stern, H., and Rubin, D. (1995). *Bayesian Data Analysis*. London: Chapman and Hall. 13

Holland, P. (1986). "Statistics and Causal Inference." *Journal of American Statistical Association*, 81: 945–960. 2, 5, 9

Katz, J., P., C., Dominici, F., and Zeger, S. (2005). "Treatment Effects of Maternal Micronutrient Supplementation vary by Percentiles of the Birth Weight Distribution in rural Nepal Distribution in Rural Nepal." Technical report, Department of Biostatistics Johns Hopkins University. 5, 8

Katz, J., West, J. J., Khatry, S., Pradhan, E., and LeClerq, S. (2000a). "Low-dose vitamin A or beta-carotene supplementation does not reduce early infant mortality: a double masked, randomized controlled community trial in Nepal." *American Journal of Clinical Nutrition*, 71: 1570–1576. 3

Katz, J., West, K., Khatry, S., LeClerq, S., Christian, P., Pradhan, E., and Shrestha, S. (2001). "Twinning rates and survival of twins in rural Nepal." *International Journal of Epidemiology*, 30: 802–7. 13

Katz, J., West, K., Khatry, S., Pradhan, E., LeClerq, S., Christian, P., Wu, L., Adhikari, R., Shrestha, S., and Sommer, A. (2000b). "Maternal low-dose vitamin A or beta-carotene supplementation has no effect on fetal loss and early infant mortality: a randomized cluster trial in Nepal." *American Journal of Clinical Nutrition*, 71: 1570–6. 13

Lechtig, A., Yarbrough, C., Delgado, H., Habicht, J., Marorelli, R., and Klein, R. (1975). "Influence of maternal nutrition on birth weight." *American Journal of Clinical Nutrition*, 28: 1223–1233. 3

Liang, K.-Y., Zeger, S., and Qaqish, B. (1992). "Multivariate regression analyses for categorical data (with discussion)." *Journal of the RoyaL Statistical Society, B*, 54: 3–40. 9

McIntire, D., Bloom, S., Casey, B., and Leveno, K. (2001). "Birth weight in relation to morbidity and mortality among newborn infants." *New England Journal of Medecine*, 340: 1234–1238. 3

Rahmathullah, L., Tielsch, J., Thulasiraj, R., Katz, J., Coles, C., Devi, S., John, R., Sadanand, A., and Edwin, K. (2003). "Impact of Newborn Vitamin A Dosing on Early Infant Mortality: A Community-Based Randomized Trial in South India." *British Medical Journal*, 327: 254–7. 13

Rasmussen, K. M. (2001). "Is there a causal relationship between iron deficiency anemia and weight at birth, length of gestation and perinatal mortality?" *Journal of Nutrition*, 131: 590–603S. 3

Rubin, D. B. (1978). "Bayesian Inference for Causal Effects: The Role of Randomization." *The Annals of Statistics*, 6: 34–58. 2, 5, 9

— (1987). *Imputation for no-responses in surveys*. New York: Wiley. 2, 5, 7, 8

Tanner, M. A. (1991). *Tools for Statistical Inference – Observed Data and Data Augmentation Methods*, volume 67 of *Lecture Notes in Statistics*. New York: Springer-Verlag. 2, 5, 13, 17

Tanner, M. A. and Wong, W. H. (1987). "The calculation of posterior distributions by data augmentation." *Journal of the American Statistical Association*, 82(398): 528–550. 2, 5, 17

Tierney, L. (1994). "Markov chains for exploring posterior distributions (with discussion)." *Annals of Statistics*, 22: 1701–1762. 13

West, J. J., Katz, J., Khatry, S., S.K. LeClerq, and Pradhan, E. (1999). "Double blind, clustered randomized trial of low dose supplementation with vitamin A or beta-caroten on mortality related to pregnancy in Nepal. The NNIPS-2 study group." *British Medical Journal*, 318: 570–575. 3

Table 1: Descriptive statistics: type of micronutrient supplementation, sample size (N), average birth weight; percent deaths, percent missing birth weights, percent weights measured after the 72 hours. The average birth weights are calculated based upon one imputed data set. The average birth weights obtained by excluding the babies with missing data and measured after the 72 hours are within parentheses.

Treatment	N	average bw (grams)	% missing	% deaths among the missing	% bw after 72 hours
Iron + Folate + vit A	766	2640 (2750)	7.0	34	10
Multiple + vit A	870	2654 (2784)	6.7	39	12.1
vit A	866	2573 (2714)	8.0	39	12.7

Table 2: Descriptive statistics: type of micronutrient supplementation, sample size (N), average birth weight and (standard deviation). The average birth weights are obtained by excluding the babies with missing data and by replacing the birth weights measured after the 72 hours with their predicted values at time zero.

Treatment	N	average bw (standard deviation)
Iron + Folate + vit A	766	2640 (465)
Multiple + vit A	870	2654 (458)
vit A	866	2573 (467)

Table 3: Definition of parameters of scientific interest for estimating the effects of micronutrient supplementation on birth weight and on infant mortality as a function of birth weight percentiles. The subscripts i and i' indicate two different infants.

Percentile-specific Effects on Birth Weight	
Average	$TE^W = E[W_i(1) - W_i(0)]$
p-specific	$\Delta_p^W = Q_1(p) - Q_0(p)$
Percentile-specific Effects on Infant Mortality	
Average	$TE^Y = E[Y_i(1) - Y_i(0)]$
p-specific	$\Delta_p^Y = E[Y_i(1) \mid F_1(W_i(1)) = p] - E[Y_i(0) \mid F_1(W_i(0)) = p]$
P-Stratification	$\begin{cases} \tau_1^Y = E[Y_i(1) - Y_i(0) \text{ given } W_i(0) \leq 2500 \ \& \ W_i(1) - W_i(0) \leq 50] \\ \tau_2^Y = E[Y_i(1) - Y_i(0) \text{ given } W_i(0) \leq 2500 \ \& \ W_i(1) - W_i(0) > 50] \\ \tau_3^Y = E[Y_i(1) - Y_i(0) \text{ given } W_i(0) > 2500 \ \& \ W_i(1) - W_i(0) \leq 50] \\ \tau_4^Y = E[Y_i(1) - Y_i(0) \text{ given } W_i(0) > 2500 \ \& \ W_i(1) - W_i(0) > 50] \end{cases}$

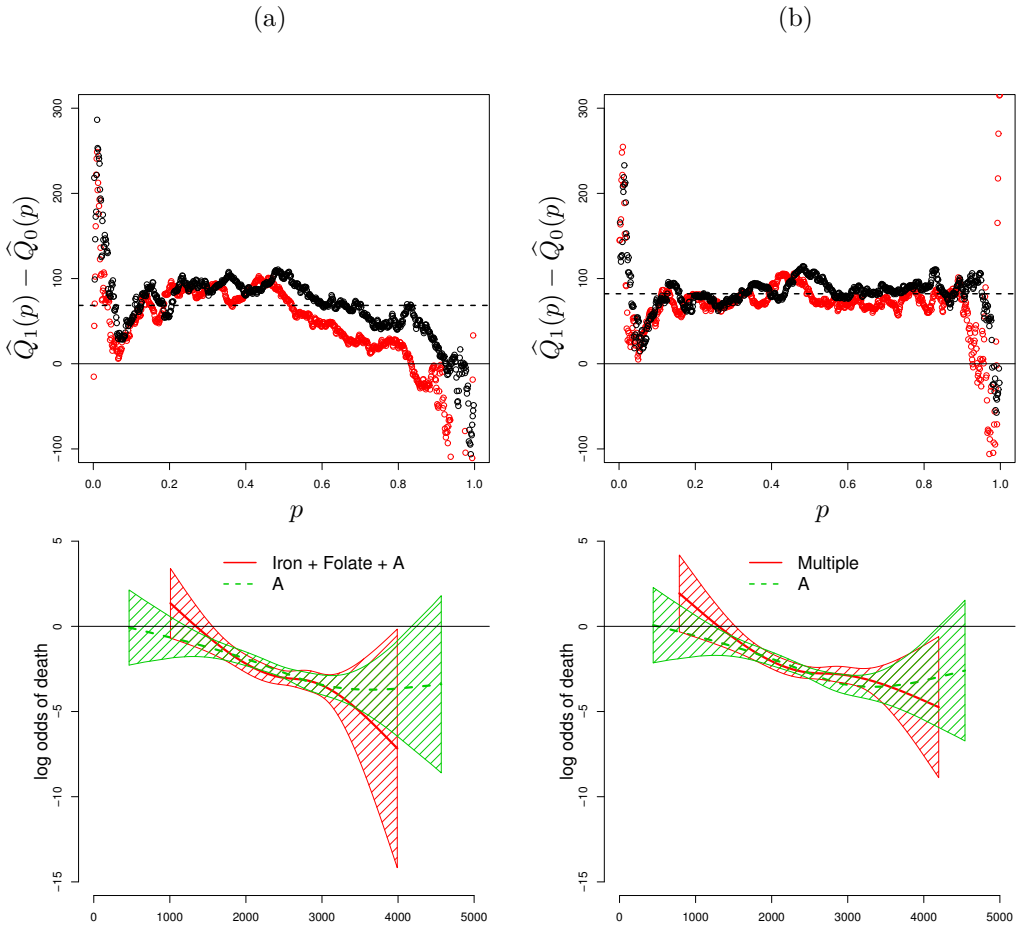


Figure 1: Top: Differences between empirical quantile functions of the birth weights for the treated and control groups. Panel (a) shows the quantile differences for the groups F+I+A versus A. Panel (b) shows the quantile differences for the groups M+A versus A. The red dots denote quantile differences of birth weights including the ones measured after the 72 hours. The black dots denote quantile differences obtained from a "working data set" where the birth weight measurements taken after the 72 hours where replaced by their predicted values at time zero (details on this prediction model are provided in Section 2). The dotted horizontal line is placed at the average difference of the birth weights between the two groups. Bottom: estimated log-odds of death as smooth function of the birth weight with 95% confidence bands and plotted in correspondence to the observed range of birth weights in the two groups.

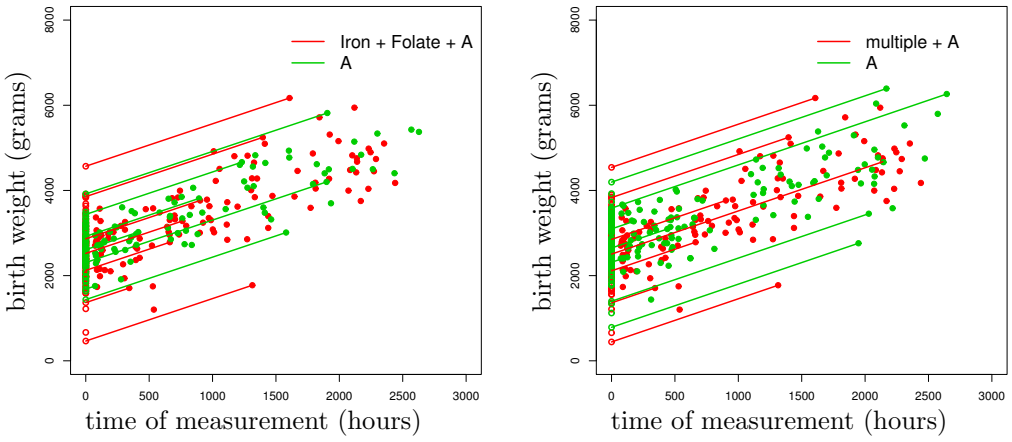


Figure 2: Birth weights plotted versus time of measurements. Red dots denote birth weights measured under the treatment and green dots denote birth weights measured under the control. The segments connect the observed measurements to the Bayesian posterior means of the predicted measurements at time zero for a random subset of the data.

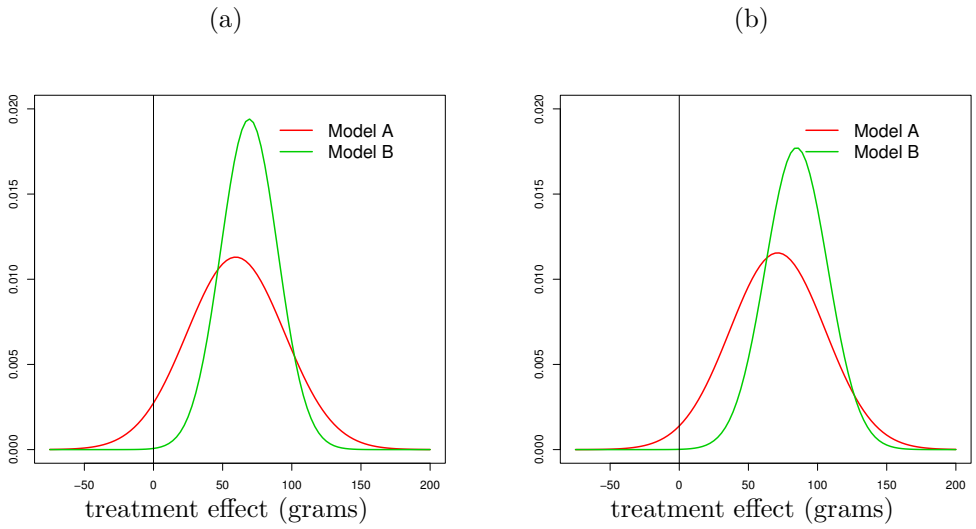


Figure 3: Marginal posterior distributions of the average treatment effect for the counterfactual model $TE^W = E(W_i(1) - W_i(0)]$ under two model specifications. Panel (a) shows the results for F+I+A compared to vit A and Panel (b) shows the results for M+A compared to vit A. The red curve denotes the posterior distribution of the average causal treatment effect obtained under a Bayesian model that accounts for the uncertainty in the imputation of the missing data, the estimation of the birth weights at time zero, and the mother's covariates (Model A). The green curve denotes the posterior distribution obtained under a Bayesian model that uses one imputed data set and that does not account for the mother's covariates (Model B).

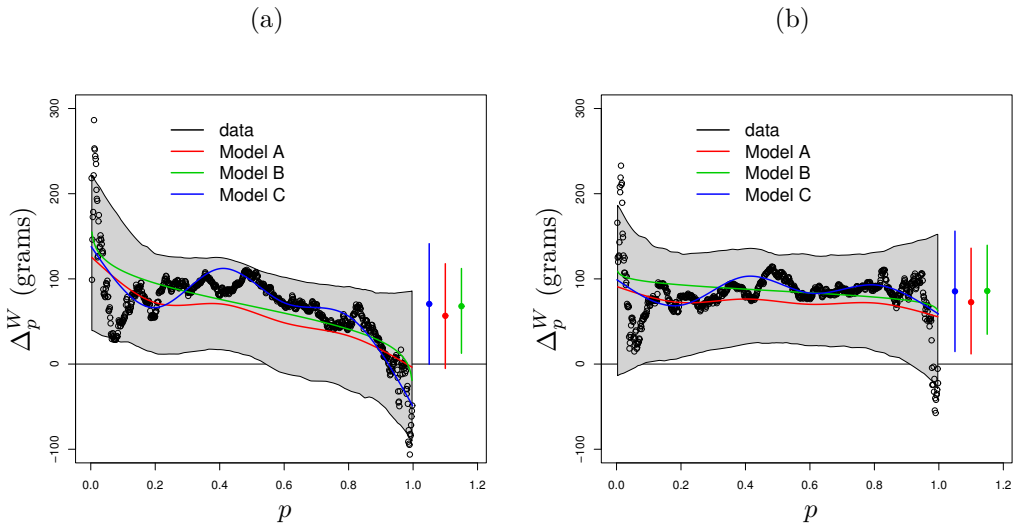


Figure 4: Marginal posterior distributions of the percentile-specific treatment effects on birth weight under Models A, B, and C denoted with red, green and blue smooth lines, respectively. The black dots are the differences in empirical quantile functions for a "working data set". The grey polygon denotes the 95% posterior confidence bands under Model A. At the far rights are shown posterior inferences and 95% uncertainty intervals of the average treatment effect for the three models.

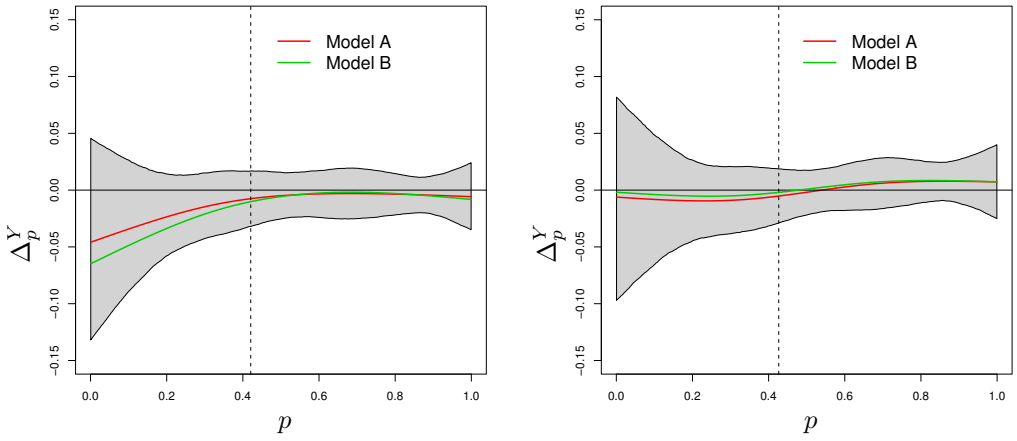
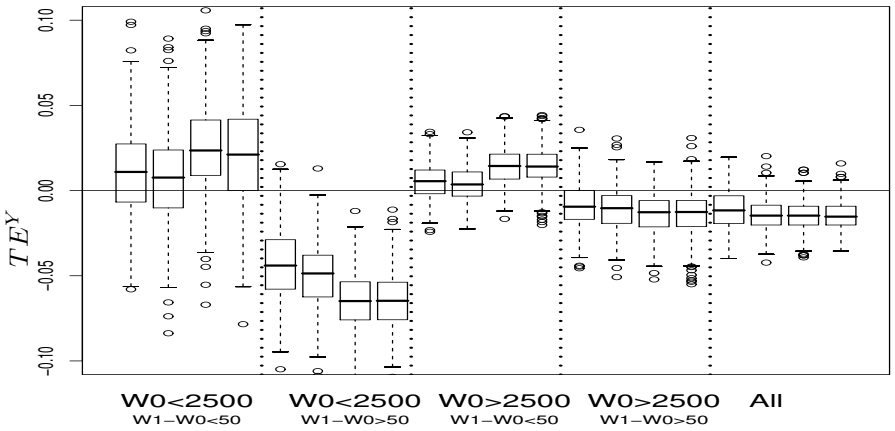


Figure 5: Posterior means and 95% posterior regions of the percentile-specific effects of treatment on mortality (Δ_p^Y) as a smooth function of the percentiles under Models A and B.

(a)



(b)

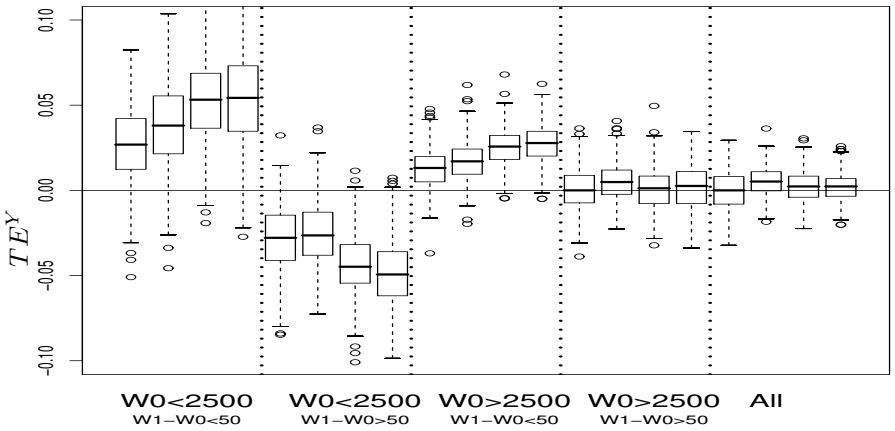


Figure 6: Posterior distributions of the average effects of treatment on mortality under Model A. Results are shown for different values of ρ and ψ. The four boxplots witin each the five sub-populations denote the posterior distribution for the following four scenarios of (ρ, ψ) : $(0.9, 1.5), (0.9, 25), (0.5, 1.5), (0.9, 25)$. The posterior distributions are shown separately for five sub-populations of infants: 1) LBW infants for whom there is an effect of treatment on birth weight smaller than 50 grams; 2) LBW infants for whom there is an effect of treatment on birth weight larger than 50 grams; 3) not-LBW for whom there is an effect of treatment on birth weight smaller than 50 grams; 4) not-LBW for whom there is an effect of treatment on birth weight larger than 50 grams; and 5) all infants.

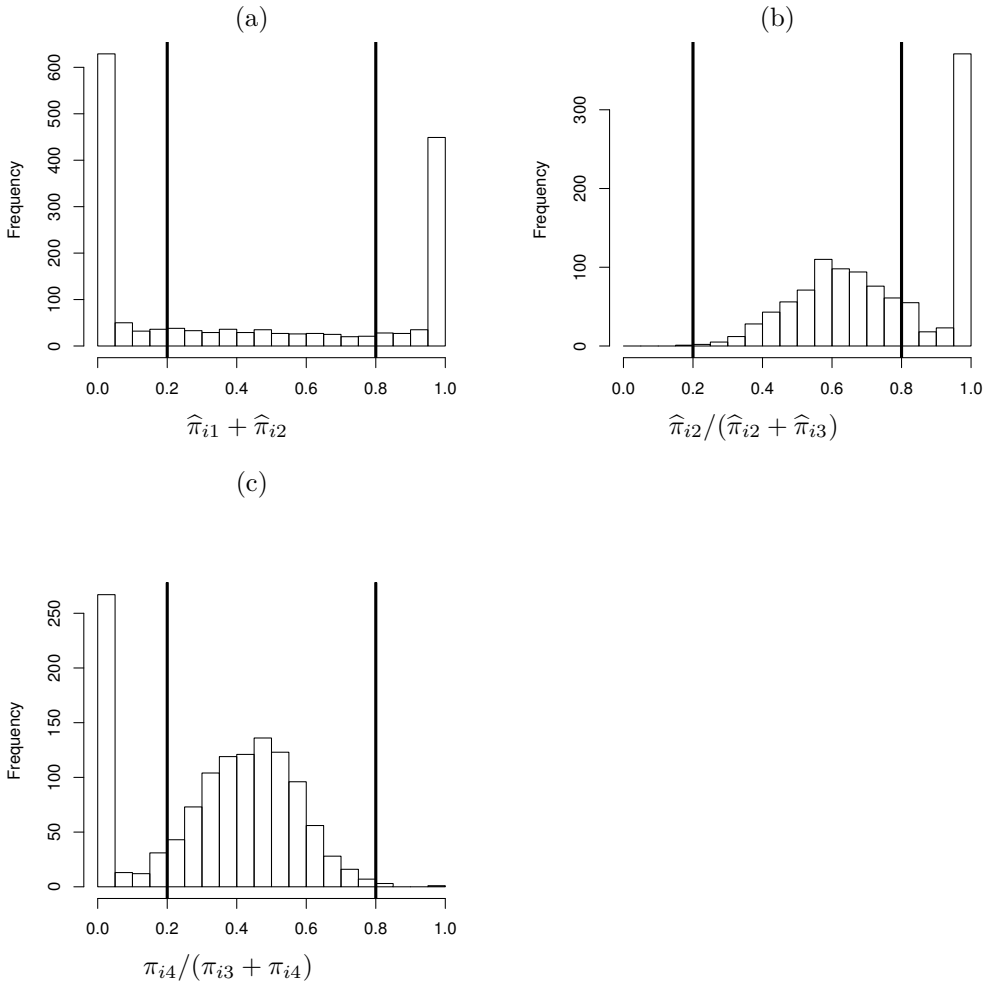


Figure 7: Panel (a): histogram of the probabilities $\widehat{\pi}_{i1} + \widehat{\pi}_{i2} = P(W_i(0) \leq 2500)$. Panel (b): histogram of the probabilities $\widehat{\pi}_{i2}/(\widehat{\pi}_{i1} + \widehat{\pi}_{i2}) = P(W_i(1) - W_i(0) > 50 \mid W_i(0) \leq 2500)$. Panel (c):histogram of the probabilities $\widehat{\pi}_{i4}/(\widehat{\pi}_{i3} + \widehat{\pi}_{i4}) = P(W_i(1) - W_i(0) > 50 \mid W_i(0) > 2500)$. The vertical lines are placed at 0.2 and 0.8.