

# Estimating percentile-specific treatment effects in counterfactual models: a case-study of micronutrient supplementation, birth weight and infant mortality

Francesca Dominici, Scott L. Zeger, Giovanni Parmigiani, Joanne Katz and Parul Christian

*Johns Hopkins University, Baltimore, USA*

[Received February 2005. Revised August 2005]

**Summary.** 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 (LBW) infants (2500 g or under), an effective intervention increases the birth weight among the smallest babies. The paper defines population and counterfactual parameters for estimating the treatment effects on birth weight and on survival as functions of the percentiles of the birth weight distribution. We use a Bayesian approach with data augmentation to approximate the posterior distributions of the parameters, taking into account uncertainty that is associated with the imputation of the counterfactuals. This approach is particularly suitable for exploring the sensitivity of the results to unverifiable modelling assumptions and other prior beliefs. We estimate that the average causal effect of the treatment on birth weight is 72 g (95% posterior regions 33–110 g) and that this causal effect is largest among the LBW infants. Posterior inferences about average causal effects of the treatment on birth weight are robust to modelling assumptions. However, inferences about causal effects for babies at the tails of the birth weight distribution can be highly sensitive to the unverifiable assumption about the correlation between the observed and the counterfactuals birth weights. Among the LBW infants who have a large causal effect of the treatment on birth weight, we estimate that a baby receiving the treatment has 5% less chance of death than if the same baby had received the control. Among the LBW infants, we found weak evidence supporting an additional beneficial effect of the treatment on mortality independent of birth weight.

**Keywords:** Causal inference; Data augmentation; Direct effects; Mediation; Percentile-specific effects; Post-treatment variables; Quantile functions

## 1. Introduction

The reduction of infant mortality remains a major public health goal (Child Health Research Project, 1996), particularly in developing countries where current rates are an order of magnitude higher than in Europe, North America and Japan. In developing countries, higher infant mortality is partially caused by poor maternal and foetal nutrition as reflected in the distribution of infant birth weights. One intervention trial has attempted to reduce infant mortality by improving maternal micronutrient sufficiency (Christian *et al.*, 2003a). The idea is to improve maternal nutritional status, thereby improving foetal growth and reducing the risk of infant mortality. Because infant mortality is greatest among low birth weight (LBW) infants (2500 g or under) and infants whose birth weights are very low (1500 g or under), it is assumed that an

*Address for correspondence:* Francesca Dominici, Department of Biostatistics, Johns Hopkins University, Baltimore, MD 21205, USA.  
E-mail: fdominic@jhsph.edu

effective intervention must increase birth weight among the smallest babies, i.e. 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 who are 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 intervention programme provided weekly iron, folic acid and vitamin A whereas the control was weekly vitamin A alone. The 1051 and 947 pregnant women who were assigned to the control and treatment delivered 866 and 766 live-born infants respectively. Details on the study designs including the rationale for the selection and exclusion of the women in the study are provided in Christian *et al.* (2003a). The team measured the birth weight within 72 h of delivery and then followed the infants for 1 year to determine whether or not they survived. However, among the 866 and 766 live-births for the control and treatment, 8% and 7% have missing birth weight and 12% and 10% were weighed after the 72 h respectively. We developed a random-effects model to impute the missing birth weights and to predict the 'weights at birth' conditional on all the available data. In this paper we report the results for one imputed data set. Details on the imputation model and a summary of the results accounting for the uncertainty in the imputation are described elsewhere (Katz *et al.*, 2005; Dominici *et al.*, 2005b). In addition treatments were randomized to 426 communities rather than to individual women. This can create some correlation between the birth weights and infant deaths within communities. It is a minor extension of the methods that are discussed in this paper to account for this clustering which turns out to be of negligible magnitude for the infant mortality outcome. To simplify the notation and exposition, we shall not address clustering here.

The interesting aspect of this study is that the investigators expect that antenatal iron–folic acid supplementation may affect birth weight and ultimately survival differently among the smaller and larger babies, i.e. they hypothesize that there could be an interaction between the treatment effect and the birth weight percentiles. Cox (1984) referred to this as the most basic form of interaction. Doksum and Sievers (1976) defined a similar form of interaction by allowing the treatment effect to vary as a function of the health response. Koenker and Bassett (1978) introduced quantile regression methods which model the quantile function of an outcome variable as a function of covariates and applied this approach to survival times where the regression parameters are allowed to depend on the quantile of interest (Koenker and Geling, 2001). Dominici *et al.* (2005a) recently introduced smooth quantile ratio estimation, a method for estimating the difference in medical expenditures between people with and without a disease as a function of the medical expenditures percentiles.

The second interesting question from this study is whether the antenatal iron–folic acid supplementation improves survival largely through its positive effect on birth weight. The hypothesis is that supplementation will improve intra-uterine growth, lowering the risk of LBW and thus increasing the chance of survival during the first year. Therefore we are interested in investigating the relative importance of different pathways for the antenatal iron–folic acid supplementation on survival. By one pathway, the intervention affects survival only throughout a change in birth weight (the so-called 'mediated effect'). A second possible pathway is that intervention affects survival above its effect on intra-uterine growth, i.e. through other mechanisms that do not involve birth weight. We refer to this pathway as a 'direct effect'. This question is synonymous with whether birth weight is a suitable 'surrogate' end point for 1-year survival.

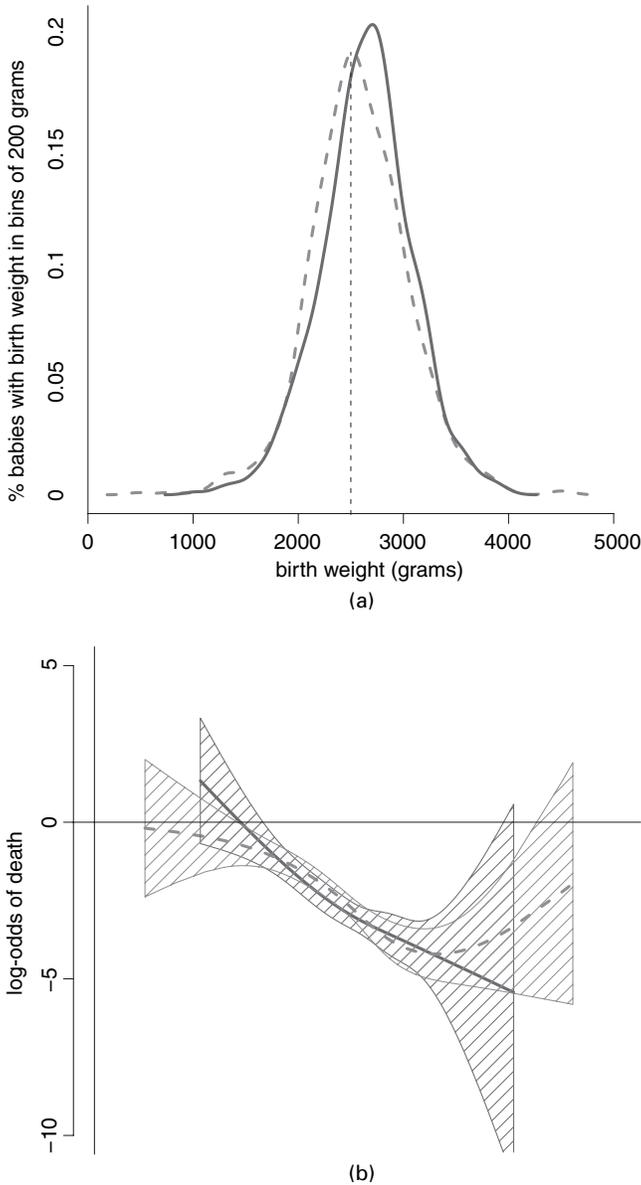
The statistical literature on surrogate end points and causal inference extensively discusses post-treatment variables in clinical trials and observational studies. Prentice (1986) first proposed criteria for a perfect surrogate (e.g. the birth weight), the most important being that the final response is conditionally independent of treatment given the surrogate. When the assumption of conditional independence is violated, related approaches have been proposed that compare results of the regression of the health response on the treatment with and without the adjustment for the intermediate variable (Freedman *et al.*, 1992; Daniels and Hughes, 1997; Buyse and Molenberghs, 1998; Begg and Leung, 2000; Leung, 2001; Molenberghs *et al.*, 2001; Xu and Zeger, 2001; Cowles, 2002). Robins (1989), Robins and Greenland (1992) and Pearl (2000) have developed identifiability results for direct and indirect causal effects under the framework of potential outcomes and they defined an ‘individual direct effect’ as the counterfactual effect of a treatment on an outcome when the intermediate variable is set at the value that it would have had if the individual had not been treated (see also Cole and Hernan (2002)). These identifiability results have been recently generalized by van der Laan and Petersen (2004). Frangakis and Rubin (2002) proposed a novel approach for defining causal effects adjusted for post-treatment variables. This approach, known as ‘principal stratification’, is based on a comparison of treatment effects on the outcome among subpopulations for whom a causal effect of treatment on the post-treatment variable did and did not occur.

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. Fig. 1(a) shows the smoothed histograms of the birth weights. Fig. 1(b) shows the estimated smooth curves with 95% confidence bands plotted in correspondence to the ranges of the measured birth weights in the two groups. These exploratory plots suggest that

- (a) the probability of death decreases as the birth weight increases and tends to rise again for the heaviest babies in the control group,
- (b) approximately 43% and 34% of the babies in the control and in the intervention groups are LBW respectively, suggesting that the treatment may reduce the percentage of LBW, and
- (c) the visual inspection of the two smoothed histograms suggests that the treatment increases birth weight for the smaller babies only, thus indicating that the treatment effect on birth weight might vary with respect to the percentiles of the birth weight distribution.

The broad objectives of this paper are to develop and apply a statistical model with counterfactual variables for this birth weight–mortality study. We shall refer to parameters from counterfactual models as ‘causal’. The contributions of this paper are

- (a) to define and compare population and causal parameters (Holland, 1986) 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),
- (b) to extend and apply a causal statistical framework to compare the causal ‘direct’ effect of the treatment on mortality, from the causal effect of the treatment on mortality that is ‘mediated’ by post-treatment changes in birth weight,
- (c) to develop a Bayesian approach with data augmentation (Tanner and Wong, 1987; Tanner, 1991; Albert and Chib, 1993; Chib and Greenberg, 1998) for approximating the marginal posterior distributions of all parameters of interest, accounting for the uncertainty about the missing counterfactuals, and
- (d) to quantify the sensitivity of causal inferences to key assumptions for which there are not direct observations in the data set.



**Fig. 1.** (a) Smoothed histograms of the birth weights for the treated (—) and control (---) groups and (b) estimated log-odds of death as a 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 (—, treatment; ---, control)

In Section 2, we introduce notation, specify our model and define the population and causal parameters. In Section 3, we define the complete likelihood function for the observed data and the missing counterfactual data. In this section, we also describe our Markov chain Monte Carlo with-data-augmentation algorithm (Tanner and Wong, 1987; Tanner, 1991) for approximating the posterior distributions of all the unknown parameters and the unobservable variables. In Section 4, we summarize the results by comparing causal and population parameter estimates.

We explore sensitivity of the causal parameter estimates to the unverifiable assumptions about counterfactuals, to model specification and to distributional assumptions. In Section 5, we discuss future research opportunities.

## 2. Definition of population and causal parameters

In this section, we define population and causal parameters that are of scientific interest in terms of counterfactual variables. To establish notation, let  $Z_i$  be the treatment indicator for live-birth  $i$  that takes values 0 or 1 to indicate the control and the treatment groups respectively. Let  $W_i^{\text{obs}}$  be the observed birth weight measurement within 72 h of the delivery, and let  $Y_i^{\text{obs}}$  be the observed mortality indicator within 1 year. Let  $n_0 = 866$  and  $n_1 = 766$  be the number of live-births for the control and the treatment groups respectively and let  $N = n_0 + n_1 = 1632$  be the total number of live-births.

Adopting a causal model with potential outcomes (Rubin, 1978; Holland, 1986), let  $\mathbf{Z}$  be the  $N$ -dimensional vector of treatment assignments with  $i$ th element  $Z_i$ , and  $W_i(\mathbf{Z})$  be the birth weight of baby  $i$  given the randomly allocated vector  $\mathbf{Z}$ . We define  $Y_i(\mathbf{Z}, \mathbf{W})$  to be the binary random variable for the mortality indicator for baby  $i$  corresponding to the vector of birth weights  $\mathbf{W}$  and the vector of treatment assignments  $\mathbf{Z}$ . We refer to  $Y_i(\mathbf{W}, \mathbf{Z})$  and  $W_i(\mathbf{Z})$  as potential outcomes. To assure a valid causal interpretation of the causal estimands defined below, we make the following usual assumptions.

- (a) *Assumption 1—stable unit treatment value assumption (Rubin, 1978)*: the potential outcomes of each baby  $i$  are unrelated to the treatment status of other babies, i.e. the birth weight and the mortality potential outcomes of each baby are not affected by the treatment assignment of others. Therefore we can write  $Y_i(\mathbf{Z}, \mathbf{W})$  and  $W_i(\mathbf{Z})$  as  $Y_i(Z_i, W_i)$  and  $W_i(Z_i)$  respectively.
- (b) *Assumption 2—ignorable assignment*: assignment to the supplementation is at random.
- (c) *Assumption 3—perfect compliance with the treatment*: all mothers take the assigned dose in both treatment groups.

Note that  $Y_i(0)$  and  $W_i(0)$  are defined for all  $N$  babies, but they are observed only 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 they are observed only for the  $n_1$  babies in the intervention group. Thus we denote the observed and the missing data as  $Y_i^{\text{obs}} = \{Y_i(z), \text{ if } z = Z_i\}$  and  $Y_i^{\text{mis}} = \{Y_i(z), \text{ if } z \neq Z_i\}$  respectively. Similar definitions apply to  $W_i^{\text{obs}}$  and  $W_i^{\text{mis}}$ .

Parameters of interest are defined in Tables 1 and 2 for birth weight and mortality respectively. The first two rows of Table 1 indicate population parameters measuring the difference between the means ( $\Delta^W$ ) and the percentiles ( $\Delta_p^W$ ) of the population of birth weights for the two treatments. Note that the parameter  $\Delta_p^W$  is defined as  $Q_1(p) - Q_0(p)$  where  $Q_1(p)$  and  $Q_0(p)$  are the quantile functions of the marginal distributions of  $W_i(1)$  and  $W_i(0)$  respectively.

The last two rows of Table 1 define causal parameters measuring the effects of the treatment on birth weight, on average ( $\tau^W$ ) and specific to the percentiles of the birth weight distribution ( $\tau_p^W$ ). Note that  $\Delta_p^W$  is a population parameter, whereas  $\tau_p^W$  is a causal parameter: in the definition of  $\Delta_p^W$ , we consider the difference in percentiles of two different distributions of birth weights. In the definition of  $\tau_p^W$ , we consider the expected difference in birth weights  $W_i(1) - W_i(0)$  for the same infant (Holland, 1986) whose control value  $W_i(0)$  is at the  $p$ -percentile of the control distribution.

Table 2 summarizes the population and causal parameters for the treatment effect on infant mortality. Before defining these parameters, we need to specify a functional relationship between

**Table 1.** Definition of population and causal parameters for estimating the effects of antenatal iron-folic acid supplementation on birth weight as a function of birth weight percentiles†

<i>Percentile-specific effects on birth weight</i>	
<i>Population parameters</i>	
Average	$\Delta^W = E[W_i(1)] - E[W_{i'}(0)] = E[W_i^{obs} Z_i = 1] - E[W_{i'}^{obs} Z_{i'} = 0]$
Percentile specific	$\Delta_p^W = E[W_i(1) F_1\{W_i(1)\} = p] - E[W_{i'}(0) F_0\{W_{i'}(0)\} = p] = Q_1(p) - Q_0(p)$
<i>Causal parameters</i>	
Average	$\tau^W(k_\rho, \psi) = E[W_i(1) - W_i(0)]$
Percentile specific	$\tau_p^W(k_\rho, \psi) = E[W_i(1) - W_i(0) F_0\{W_i(0)\} = p]$

†  $Q_1(p)$  and  $Q_0(p)$  are the quantile functions of  $W_i(1)$  and  $W_{i'}(0)$  respectively. The parameters  $k_\rho = \min_i(\rho_i)$  and  $\psi$  measure the minimum of the correlations between  $W_i(0)$  and  $W_i(1)$  and the odds ratio between  $Y_i(0)$  and  $Y_i(1)$ . The subscripts  $i$  and  $i'$  indicate two different infants.

**Table 2.** Definition of population and causal parameters for estimating the effects of antenatal iron-folic acid supplementation on infant mortality as a function of the birth weight percentiles†

<i>Percentile-specific effects on mortality</i>	
<i>Population parameters</i>	
Average	$\Delta^Y = E[Y_i(1)] - E[Y_{i'}(0)] = E[Y_i^{obs} Z_i = 1] - E[Y_{i'}^{obs} Z_{i'} = 0]$
Percentile specific	$\Delta_p^Y = E[Y_i(1) F_1\{W_i(1)\} = p] - E[Y_{i'}(0) F_0\{W_{i'}(0)\} = p]$
<i>Causal parameters</i>	
Average	$\tau^Y(k_\rho, \psi) = E[Y_i(1) - Y_i(0)]$
Percentile specific	$\tau_p^Y(k_\rho, \psi) = E[Y_i(1) - Y_i(0) F_0\{W_i(0)\} = p]$
<i>P-stratification</i>	
	$\tau_1^Y(k_\rho, \psi) = E[Y_i(1) - Y_i(0) \text{ given } W_i(0) \leq 2500 \ \& \  W_i(1) - W_i(0)  \leq 50]$
	$\tau_2^Y(k_\rho, \psi) = E[Y_i(1) - Y_i(0) \text{ given } W_i(0) \leq 2500 \ \& \  W_i(1) - W_i(0)  > 50]$
	$\tau_3^Y(k_\rho, \psi) = E[Y_i(1) - Y_i(0) \text{ given } W_i(0) > 2500 \ \& \  W_i(1) - W_i(0)  \leq 50]$
	$\tau_4^Y(k_\rho, \psi) = E[Y_i(1) - Y_i(0) \text{ given } W_i(0) > 2500 \ \& \  W_i(1) - W_i(0)  > 50]$

† The parameters  $k_\rho = \min_i(\rho_i)$  and  $\psi$  measure the minimum of the correlations between  $W_i(0)$  and  $W_i(1)$  and the odds ratio between  $Y_i(0)$  and  $Y_i(1)$ . The subscripts  $i$  and  $i'$  indicate two different infants.

infant mortality and birth weight. Substantive knowledge and our exploratory analyses indicate that the following logistic regression model is a reasonable approximation to the actual mortality process:

$$\text{logit}[\text{Pr}\{Y_i(Z_i) = 1|Z_i, W_i(Z_i)\}] = \beta_0 + \beta_1 Z_i + s\{W_i(Z_i), 3\}, \quad Z_i = 0, 1, \quad (1)$$

where  $s(\cdot)$  denotes a natural cubic spline with three knots.

By specifying this parametric model we make two key assumptions.

- (a) *Assumption 4—conditional independence of survival from the counterfactual birth weight given the treatment assignment and the observed birth weight:* for each baby, we assume that the probability of death under the treatment depends only on the birth weight under that treatment, and it does not depend on what the birth weight would have been if the

same baby had been randomized to the other group, i.e. we assume that

$$\Pr\{Y_i(Z_i) = 1|Z_i, W_i(Z_i), W_i(1 - Z_i)\} = \Pr\{Y_i(Z_i) = 1|Z_i, W_i(Z_i)\}.$$

- (b) *Assumption 5—no interaction between the direct treatment effect on survival and the birth weight:* we assume that the direct effect of the treatment on mortality is the same for all babies and does not vary with respect to the birth weight distribution, i.e. we can write

$$\text{logit}[\Pr\{Y_i(1) = 1|Z_i = 1, W_i(1) = w\}] - \text{logit}[\Pr\{Y_i(0) = 1|Z_i = 0, W_i(0) = w\}] = \beta_1.$$

This assumption can be relaxed by assuming a linear or non-linear interaction between the treatment and the birth weight, e.g. by replacing  $\beta_1 Z_i$  with  $\beta_1 \{Z_i \times W_i(Z_i)\}$ —or more generally with  $Z_i \times s_2\{W_i(Z_i), 3\}$ —in model (1).

The first two rows of Table 2 indicate population parameters measuring treatment effect on mortality, on average ( $\Delta^Y$ ) and conditional on a specific percentile of the birth weight distribution ( $\Delta_p^Y$ ). Note that  $\Delta_p^Y$  defines 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. Thus  $\Delta_p^Y$  is not a causal parameter, because these differences correspond to two different subpopulations of babies.

The last two rows of Table 2 indicate the causal parameters measuring the effects of treatment on infant mortality, on average ( $\tau^Y$ ) and specific to the percentiles of the birth weight distribution ( $\tau_p^Y$ ). Thus, for a specific  $p$ ,  $\tau_p^Y$  can be interpreted as a causal effect which compares the probability of death for the same baby  $i$  given that the assumption that his or her birth weight under the control ( $W_i(0)$ ) is at the  $p$ th percentile.

In the last row of Table 2, we use 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,  $\tau_1^Y$  and  $\tau_2^Y$  are the effects of treatment on mortality in the two subpopulations of LBW babies for whom the treatment effect on birth weight was smaller and larger than 50 g respectively. Thus a comparison between  $\tau_1^Y$  and  $\tau_2^Y$  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 babies. The parameters  $\tau_3^Y$  and  $\tau_4^Y$  are the analogues of  $\tau_1^Y$  and  $\tau_2^Y$  for the not-LBW infants, i.e. for the infants with birth weight larger than 2500 g.

All causal parameters ( $\tau$ ) depend on unverifiable assumptions about the joint distribution of the counterfactual pairs of variables  $\{W_i(0)$  and  $W_i(1)\}$  and  $\{Y_i(0)$  and  $Y_i(1)\}$ . To estimate the average causal effects ( $\tau_s$ ), we make the following key but unverifiable assumptions about the correlation between the observed outcomes and their counterfactuals.

- (a) *Assumption 6—correlation between the observed and the counterfactual birth weight:* we assume that the correlation between  $W_i(Z_i)$  and  $W_i(1 - Z_i)$ , which is denoted by  $\rho_i$ , depends on the treatment effect and it is allowed to vary between 1 and  $k_\rho$ , where  $k_\rho$  is a prespecified value. More specifically, we assume that  $\rho_i = \exp(-\alpha|\Delta_{p_i}^W|)$  where  $p_i$  is the percentile of the observed birth weight of baby  $i$  and  $\alpha = -\log(k_\rho)/\max_p(\Delta_p^W)$ . Under this assumption, when  $\Delta_{p_i}^W = 0$ , then  $\rho_i = 1$ ,  $W_i(1) = W_i(0)$  and  $\tau_{p_i}^W = 0$ . In contrast, when  $|\Delta_{p_i}^W| > 0$ , we assume that the variability in the imputation of  $W_i(0)|W_i(1)$  (or  $W_i(1)|W_i(0)$ ), which is measured by  $\rho_i$ , increases as the percentile-specific treatment effect increases. As detailed below, we assume that  $k_\rho = \min_i(\rho_i)$  is equal to 0.4 and 0.9.
- (b) *Assumption 7—odds ratio between the observed and counterfactual mortality given birth weight:* let  $\mu_i(11)$  be the joint probability that the same baby  $i$  would die in both groups

defined as  $P\{Y_i(Z_i) = 1, Y_i(1 - Z_i) = 1 | W_i(Z_i), W_i(1 - Z_i)\}$ . As detailed below, we assume that the odds ratio  $\psi = \mu(11)\mu(00)/\mu(10)\mu(01)$  is equal to 1.5 and 25.

Neither parameters  $k_\rho$  and  $\psi$  are identified by the data and need to be prespecified. As a guide for reasonable choices of  $k_\rho$ , we have used data from this randomized trial and from other data sources (Rahmathullah *et al.*, 2003; Katz *et al.*, 2000b, 2001) to estimate the correlations of birth weights for two successive children who were born to the same mother and birth weights for twins. We found that these correlations range from 0.45 to 0.7. Therefore we set  $k_\rho = 0.4$  and  $k_\rho = 0.9$  so that we allow  $\rho_i$  to vary between 1, when the treatment effect is zero, and 0.9 (or 0.4), when the treatment effect achieves its maximum.

The analogous odds ratios for mortality were estimated to be 1.8 and 52 respectively. For  $\psi$ , we use the values 1.5 and 25. We think that choosing 25 as an upper bound rather than the twin value 52 is reasonable. The rationale is that twins share a womb and compete for resources. They therefore tend to be smaller and at higher risk of death. In Section 4, we shall study the dependence of our causal inferences on our choices for  $k_\rho$  and  $\psi$ .

### 3. A Bayesian implementation of ‘causal’ inference

In this section, we define a Bayesian approach for approximating the marginal posterior distributions of the population and of the ‘causal parameters’ (e.g. the parameters in the counterfactual models) that were defined in Section 2. We start by defining the likelihood function for the complete data as

$$L(\eta_1, \eta_2) = \prod_{i=1}^N \Pr\{Y_i(1), Y_i(0) | W_i(1), W_i(0), \eta_1\} f\{W_i(1), W_i(0) | \eta_2\}. \tag{2}$$

In Section 3.1, we specify  $f\{W_i(1), W_i(0) | \eta_2\}$  as a mixture of normal distributions. In Section 3.2, we specify an odds ratio association model for bivariate binary variables  $P\{Y_i(1), Y_i(0) | W_i(1), W_i(0), \eta_1\}$  (Liang *et al.*, 1992). This model will be consistent with equation (1). In Section 3.3, we then detail the elicitation of the prior distributions and the implementation of the Markov chain Monte Carlo methods with data augmentation to obtain posterior samples of all the unknown parameters and the missing counterfactuals variables.

#### 3.1. Statistical model for birth weight

We begin our specification of the joint distribution in equation (2), by assuming that the marginal distributions of the random variables  $W_i(z)$ ,  $z = 0, 1$ ,  $i = 1, \dots, N$ , are a mixture of  $J$  ( $=3$ ) normal distributions:

$$f_z\{W_i(z) | \mu_z, \sigma_z^2, \gamma_z\} = \prod_{j=1}^J \gamma_{zj} \phi\{W_i(z); \mu_{zj}, \sigma_{zj}^2\}, \quad z = 0, 1, \tag{3}$$

where  $\phi(x; \mu, \sigma^2)$  is the density of a normal distribution with mean  $\mu$  and variance  $\sigma^2$ ,  $\mu_z = (\mu_{1z}, \mu_{2z}, \mu_{3z})$ ,  $\sigma_z = (\sigma_{1z}, \sigma_{2z}, \sigma_{3z})$  and  $\gamma_z = (\gamma_{1z}, \gamma_{2z}, \gamma_{3z})$ , where  $\gamma_{jz}$  are the mixing probabilities with  $\sum_{j=1}^J \gamma_{jz} = 1$ . To identify the mixture we set the constraint  $\mu_{1z} < \mu_{2z} < \mu_{3z}$  (Kadane, 1975). We further assume that  $\sigma_{1z} = \sigma_{3z} = \sigma_{2z}\sqrt{2}$ : assigning a larger variance to the outside components of the mixture is designed to capture heavy-tailed distributions while achieving greater parsimony. For ease of notation, we shall set  $\sigma_z = \sigma_{2z}$ . These choices were guided by exploratory analyses which used a clustering algorithm for parameterized Gaussian mixtures (Fraley and Raftery, 2002). This algorithm estimates the number of the components in a mixture and the corresponding parameters.

This distributional assumption allows the parameters  $\Delta_p^W$  and  $\tau_p^W$  to vary flexibly as functions of the percentiles  $p$  of the birth weight distribution. If instead of the mixture model (3) we assumed that  $W_i(z) \sim N(\mu_z, \sigma_z)$ , then

$$\Delta_p^W = \mu_1 - \mu_0 + (\sigma_1 - \sigma_0) \Phi^{-1}(p).$$

Therefore, the simpler assumption of normality for the marginal distributions of  $W_i(0)$  and  $W_i(1)$  imposes a specific parametric form for  $\Delta_p^W$  which does not depend on  $p$  for  $\sigma_1 = \sigma_0$ . In Section 4, we shall calculate the posterior probability of  $\sigma_0 \neq \sigma_1$  to provide evidence in favour of the assumption that  $\Delta_p^W$  depends on  $p$ , and we shall explore the sensitivity of the posterior distribution of  $\Delta_p^W$  as a function of  $p$ , under the mixture model, under the assumption of normality with unequal variances and under the simpler assumption of normality with  $\sigma_0 = \sigma_1$ .

To allow for a correlation between  $W_i(0)$  and  $W_i(1)$ , we assume that the standardized variables  $\Phi^{-1}[F_z\{W_i(z)\}]$ ,  $z = 0, 1$ , have a bivariate normal distribution with mean 0, variance 1 and correlation  $\rho_i$ , where  $\Phi$  is the cumulative distribution function of a standard normal distribution and  $F_z$  is the cumulative distribution function of  $W_i(z)$ . As stated in assumption 6, we assume that the parameters  $\rho_i$  depend on the treatment effect:

$$\rho_i = \text{corr}(\Phi^{-1}[F_0\{W_i(0)\}], \Phi^{-1}[F_1\{W_i(1)\}]) = \exp(-\alpha|\Delta_{\rho_i}^W|).$$

To implement this approach, we replace  $\Delta_p^W$  with its estimate obtained by smoothing across percentiles the empirical quantile functions.

In this formulation for the joint distribution of  $(W_i(0), W_i(1))$ , letting  $\rho_i = 1$  for all babies corresponds to the rank preservation assumption that was used by Efron and Feldman (1991). Our specification allows for an interpretable parameter  $\rho_i$  for each baby capturing the correlation between  $W_i(0)$  and  $W_i(1)$ , while allowing for a flexible representation of the two marginal distributions. An alternative stochastic generalization of the rank preservation assumption, obtained by specifying a probabilistic distribution on the ranks, has also been developed by Dobbin and Louis (2003).

### 3.2. Statistical model for infant mortality given birth weight

We specify the counterfactual model for the joint distribution of the two outcome indicators conditional on the birth weights. Following Liang *et al.* (1992), we parameterize the  $2 \times 2$  joint distribution  $[Y_i(0), Y_i(1)|W_i(0), W_i(1)]$  in terms of the margins and the odds ratio. Specifically, we assume that

$$\begin{aligned} P\{Y_i(0) = y_i(0), Y_i(1) = y_i(1)|W_i(0), W_i(1), \eta_1\} \\ = \mu_i(0)^{y_i(0)} \{1 - \mu_i(0)\}^{1-y_i(0)} \mu_i(1)^{y_i(1)} \{1 - \mu_i(1)\}^{1-y_i(1)} \\ + (-1)^{y_i(0)-y_i(1)} \{\mu_i(11) - \mu_i(0) \mu_i(1)\} \end{aligned} \tag{4}$$

where  $\mu_i(1) = \Pr\{Y_i(Z_i) = 1|Z_i, W_i(Z_i)\}$  is defined in equation (1). The parameter  $\mu_i(11) = \Pr\{Y_i(0) = Y_i(1) = 1|W_i(0), W_i(1)\}$  is a known function of the marginal probabilities  $\mu_i(1)$  and  $\mu_i(0)$  and of the prespecified odds ratio  $\psi$ .

### 3.3. Prior distributions and computation

The distributional assumptions in Sections 3.1 and 3.2 involve the following vectors of unknown parameters:

**Table 3.** Prior distributions on the unknown parameters of the mixture†

Parameter	Prior distribution
$\beta$	Flat
$\mu_0$	$N_3\{(1500, 2500, 3500), 500^2 I\}$
$\mu_1$	$N_3\{(2000, 3000, 3500), 500^2 I\}$
$\sigma_0^2$	$LN\{\log(400^2), 0.8\}$
$\sigma_1^2$	$LN\{\log(400^2), 0.8\}$
$\gamma_0$	Dirichlet $(10, \frac{1}{3}, \frac{1}{3}, \frac{1}{3})$
$\gamma_1$	Dirichlet $(10, \frac{1}{3}, \frac{1}{3}, \frac{1}{3})$

† $I$  denotes a  $3 \times 3$  identity matrix,  $500^2$  denotes the prior variance, LN denotes the log-normal distribution with prior mean 400 and prior standard deviation 0.8.

- (a)  $\eta_1 = (\beta, \psi)$  where  $\beta$  includes  $\beta_0, \beta_1$  and the spline coefficients that are defined in the regression model (1);
- (b)  $\eta_2 = (\mu_0, \mu_1, \sigma_0, \sigma_1, \gamma_0, \gamma_1, k_\rho)$ , denoting all the unknown parameters of the mixture (3).

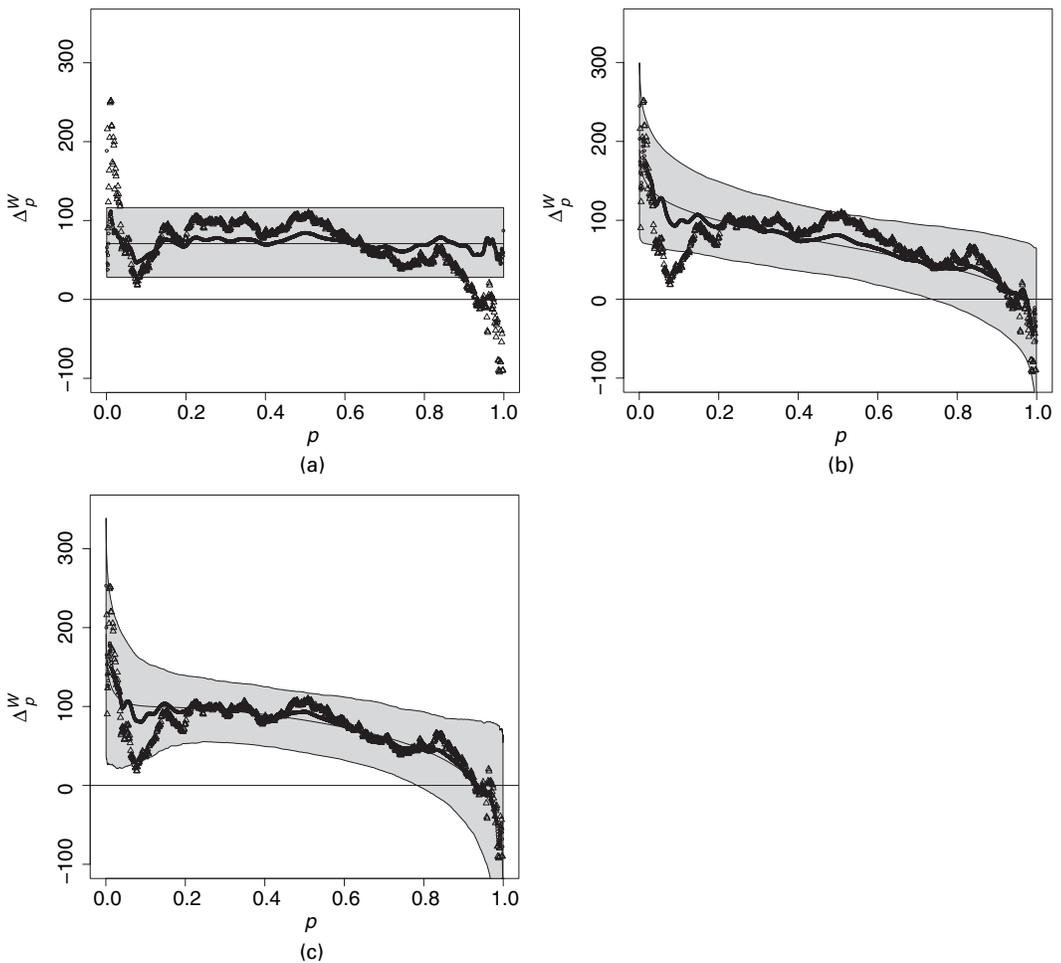
As stated in assumptions 6 and 7, the parameters  $k_\rho$  and  $\psi$  measure the association between the observed outcomes and their counterfactuals and they cannot be identified from the observed data. Thus,  $k_\rho$  and  $\psi$  are prespecified fixed values. In Section 4 we shall present the results for  $(k_\rho, \psi) = (0.4, 1.5), (0.4, 25), (0.9, 1.5), (0.9, 25)$ . We specify prior distributions on all the other parameters of the mixture that are proper but sufficiently vague to achieve goodness of fit to the observed birth weights. These choices are summarized in Table 3.

To investigate the posterior distributions of all the parameters of interest we implement Markov chain Monte Carlo methods with data augmentation for imputing the missing data (Tanner, 1991; Gelman *et al.*, 1995). Bayesian sampling of parameters of normal mixture distributions is typically handled by introducing auxiliary variables representing mixture component indicators, which results in closed form full conditionals (Diebolt and Robert, 1994). In our case, this option was not practical because of the special correlation structure that we used, and because the unobserved birth weight variables enter the logistic component of the likelihood as well. We thus implemented a Metropolis-within-Gibbs (Tierney, 1994) approach, in which both the parameters and the counterfactual variables are sampled by using a random-walk proposal, truncated to the region that is defined by the constraints wherever applicable.

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  $\Delta_p^W$ , we sort  $W_i(0)$  and  $W_{i'}(1)$  within the two groups of treated and untreated babies separately and then take their difference. To calculate a posterior sample of  $\tau_p^W$ , we sort by  $W_i(0)$  and then take the difference between the sorted  $W_i(0)$  and its matched  $W_i(1)$  for the same infant  $i$ . To calculate a posterior sample of  $\Delta_p^Y$ , we first sort  $Y_i(0)$  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. Finally, to calculate a posterior sample of  $\tau_p^Y$ , we sort  $Y_i(0)$  with respect to  $W_i(0)$ , and then we take the difference between the sorted  $Y_i(0)$  and its matched  $Y_i(1)$  for the same baby  $i$ . We smoothed the posterior samples of these percentile-specific parameters to reduce Monte Carlo variability in the posterior probability bounds.

4. Results

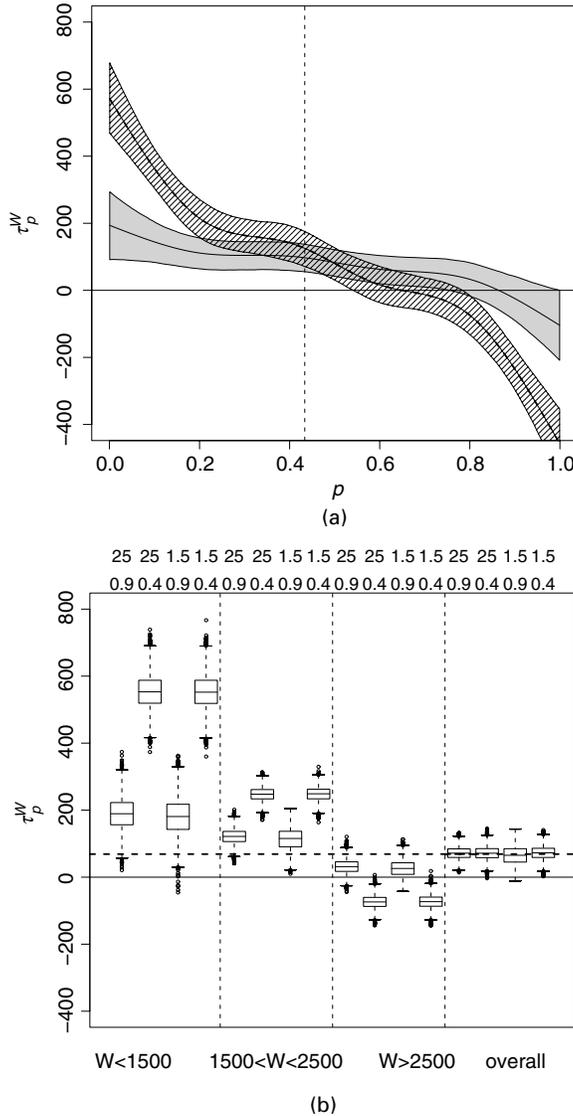
Fig. 2 shows the posterior means and 95% posterior regions for the population percentile-specific treatment differences in birth weight ( $\Delta_p^W$ ) under the three modelling assumptions for  $(W_i(0), W_i(1))$ . In Fig. 2(a)  $(W_i(0), W_i(1))$  is assumed to have a bivariate normal distribution with equal variances. In Fig. 2(b) it has bivariate normal distribution with unequal variances. In Fig. 2(c) it has a mixture of normal distributions with correlation  $\rho_i$  as defined in Section 2. The triangles denote the difference between the empirical quantile functions for the observed data. The dots denote the differences between the posterior means of the birth weights for the two treatment groups. The lines denote the posterior means and 95% posterior regions of the theoretical functions  $Q_1(p) - Q_0(p)$ . These theoretical functions are equal to  $\mu_1 - \mu_0$  in Fig. 2(a),  $\mu_1 - \mu_0 + (\sigma_1 - \sigma_0) \Phi^{-1}(p)$  in Fig. 2(b) and  $F_1^{-1}(w; \theta_1) - F_0^{-1}(w; \theta_0)$  where  $F_0$  and



**Fig. 2.** Posterior means and 95% posterior regions of the percentile-specific effects of treatment on birth weight ( $\Delta_p^W$ ) under the following modelling assumptions for  $(W_i(0), W_i(1))$ : (a) bivariate normal distribution with equal variances; (b) bivariate normal distribution with unequal variances; (c) mixture of three normal distributions with correlation  $\rho_i$  as defined in Section 2 (results are shown for  $\min_i(\rho_i) = k_\rho = 0.9$  and  $\psi = 25$ )

$F_1$  are the cumulative distribution functions of the mixture normal distributions in Fig. 2(c) respectively.

Under the three modelling assumptions for the birth weights, the difference in posterior means for the two treatment groups (dots) are roughly the same and are close to the observed difference (triangles). However, the percentile-specific treatment effects differ substantially from one another in the tails. In Fig. 2(a),  $\Delta_p^W$  is a constant function of  $p$  as is reflected in the flat line relationship. If we fit a bivariate normal distribution without the constraint of equal variances as is done in Fig. 2(b), the treatment effect is estimated to decrease from more than 100 g in the



**Fig. 3.** (a) Posterior means and 95% posterior regions of the percentile-specific causal effect of treatment on birth weight ( $\tau_p^W$ ) for  $k_\rho = 0.4$  (▨) and  $k_\rho = 0.9$  (▢) and (b) sensitivity analysis of the posterior distributions of the causal effect of treatment on birth weight ( $\tau_p^W$ ) separately for three subpopulations of babies  $W_i(0) < 1500$  g,  $1500 < W_i(0) \leq 2500$  g and  $W_i(0) > 2500$  g and overall for all babies



Fig. 3 (a) shows the posterior means and 95% posterior regions of the  $p$ -specific causal effects of treatment on birth weight ( $\tau_p^W$ ) under the mixture model, for  $k_\rho = 0.9$  (shaded area) and  $k_\rho = 0.4$  (hatched area). The vertical line is placed at the 0.42-percentile corresponding to 2500 g in the control sample. Note that, when  $k_\rho$  becomes smaller, then population and causal parameter inferences differ little in the middle of the distribution, but substantially towards the tails of the birth weight distribution. Among LBW infants, we found that the average causal effects of the intervention on the birth weight are equal to 143 g (95% posterior regions, 83–270 g) and to 337 g (95% posterior regions, 211–631 g) for  $k_\rho = 0.9$  and  $k_\rho = 0.4$  respectively.

Fig. 3(b) shows the sensitivity of the posterior distributions of the causal effect of treatment on birth weight ( $\tau_p^W$ ) separately for three subpopulations of babies ( $W_i(0) \leq 1500$ ,  $1500 < W_i(0) \leq 2500$  and  $W_i(0) < 2500$ ), and overall for all babies, with respect to  $(k_\rho, \psi)$ . The horizontal broken line is placed at the sample mean difference ( $\Delta^W$ ). The two rows on top of Fig. 3(b) denote the values of  $\psi$  (top row) and  $k_\rho$  (bottom row) that were used in the analysis. Within each subpopulation, these causal effects are very sensitive to  $k_\rho$  but not to  $\psi$ . However, the average causal effect of supplementation on birth weight ( $\tau^W$ )—which was estimated to be 72 g (95% posterior regions, 33–110 g)—is robust to the values that were assigned for  $k_\rho$  and  $\psi$ .

Fig. 4(a) shows the posterior means and 95% posterior regions of the percentile-specific difference in infant mortality rates between the treatment and control populations ( $\Delta_p^Y$ ) plotted with respect to the percentiles of the birth weight distributions. For a specific  $p$ ,  $\Delta_p^Y$  is the difference in the probability of death between the babies with birth weights  $W_i(1)$  and  $W_i(0)$ , 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 g in the control sample. 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.

Fig. 4(b) shows the posterior means and 95% posterior regions of the  $p$ -specific causal effect of the treatment on infant mortality ( $\tau_p^Y$ ) plotted with respect to the percentiles of  $W_i(0)$  for  $k_\rho = 0.9$  and  $k_\rho = 0.4$  (bold curve). For a specific  $p$ ,  $\tau_p^Y$  is defined as the difference in the probability of death for the same baby  $i$  whose control birth weight  $\{W_i(0)\}$  is at the  $p$ th percentile. For  $k_\rho = 0.9$ , we found that the causal effect of supplementation on mortality adjusted by birth weight is negative (intervention is better) for the smaller babies and that this effect diminished for the larger babies, although with wide posterior regions. Posterior inferences on  $\tau_p^Y$  are quantitatively but not qualitatively sensitive to  $k_\rho$  at the smallest percentiles. For  $k_\rho = 0.4$  (bold curve), there is stronger support for a beneficial ‘direct’ effect of the supplementation on mortality among the very small babies only.

Fig. 4(c) shows the estimated posterior distributions of the causal effects of treatment on mortality for different values of  $(k_\rho, \psi)$  among four subpopulations of babies:

- (a) LBW infants for whom there is a causal effect of treatment on birth weight that is smaller than 50 g ( $\tau_1^Y$ );
- (b) LBW infants for whom there is a causal effect of treatment on birth weight that is larger than 50 g ( $\tau_2^Y$ );
- (c) not-LBW infants for whom there is a causal effect of treatment on birth weight that is smaller than 50 g ( $\tau_3^Y$ );
- (d) not-LBW infants for whom there is a causal effect of treatment on birth weight that is larger than 50 g ( $\tau_4^Y$ ).

The four box plots at the far right-hand side show the estimated posterior distributions of the total (direct plus mediated) causal effect of supplementation on mortality on average for all babies ( $\tau^Y$ ).

The four box plots on the left-hand side (posterior distributions of  $\tau_1^Y$ ) indicate that, among the LBW babies with little change in birth weight after the supplementation, there is only weak evidence that antenatal iron–folic acid supplementation affects survival. The second set of four box plots (posterior distributions of  $\tau_2^Y$ ) suggest that, among the LBW babies with absolute changes in birth weight after the supplementation larger than 50 g, there is much stronger evidence that the antenatal iron–folic acid supplementation affects survival. The posterior means of these ‘mediated’ causal effects for  $k_\rho = 0.9$  and  $k_\rho = 0.4$  are equal to  $-0.04$  and  $-0.052$  (95% posterior regions,  $-0.077$ – $0$  and  $-0.085$ – $0.004$ ) respectively. These results indicate that an LBW infant receiving the intervention has a roughly 5% smaller chance of death than if the same baby had received the control intervention. This higher chance of death is due to changes in birth weight from the control to the treatment larger than 50 g. The posterior distributions of the parameters  $\tau_3^Y$  and  $\tau_4^Y$  indicate that there is no evidence of a beneficial effect for the infants who are not LBW. The average causal effect of supplementation on mortality that is shown on the right-hand side is robust to modelling assumptions and to  $(k_\rho, \psi)$ .

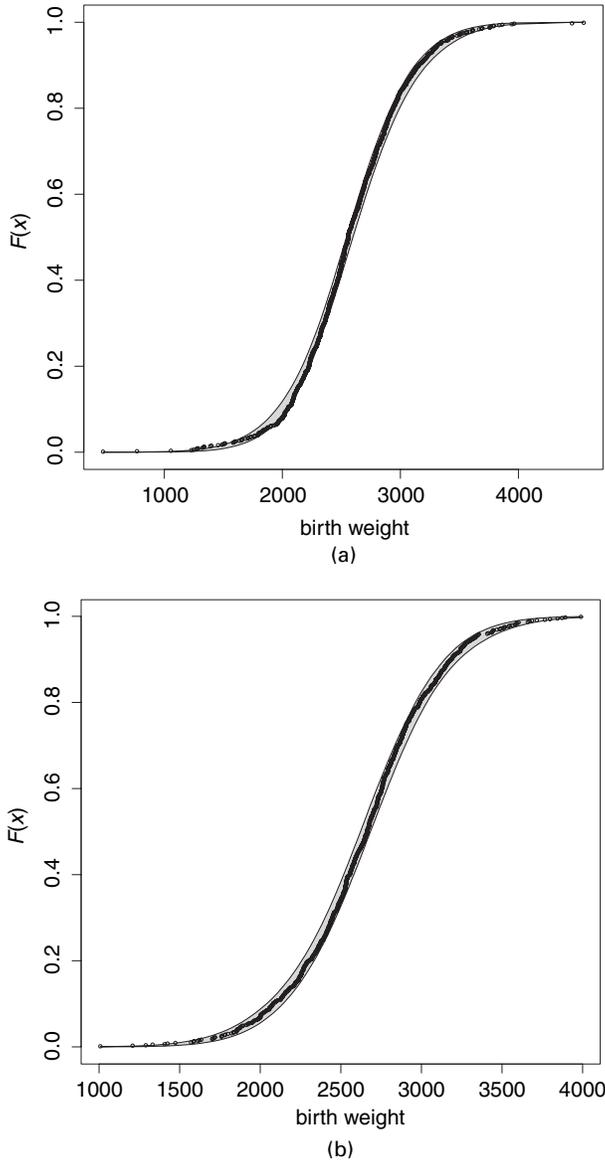
Finally, we evaluate the consistency of the model predictions and prior distributions with the patterns in the observed data. Fig. 5 shows 95% posterior regions of  $F_z\{W_i(z), \theta_z^{(j)}\}$ ,  $z = 0, 1$ , where  $F_z$  are the cumulative distribution functions from the mixture model that is defined in equation (3) and  $\theta_z^{(j)}$  are the  $j$ th posterior samples of the parameters of the mixture. The curves are the corresponding empirical cumulative distribution functions, estimated directly from the observed birth weights. We see that the model assumed is reasonably consistent with the data.

## 5. 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 LBW infants, an effective intervention must increase birth weight mainly among the smallest babies. In addition, it has been hypothesized that the supplementation could be harmful if it increases birth weight among the largest babies. A community-based trial in Nepal has shown that a multiple micronutrient supplementation increases birth weight but the limitations in the study size have to date prevented us from unambiguously establishing that this translates into a mortality benefit (Christian *et al.*, 2003b).

In this paper, we have developed 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 (mediated effect) or it improves survival even without affecting the birth weight (direct effect). Addressing these scientific questions is challenging because birth weight is a post-treatment variable (i.e. intermediate variable) that is in the causal pathway between nutritional supplementation and infant mortality.

This analysis demonstrates that inference about counterfactual treatment effects in the middle of the birth weight distribution is relatively robust to unverifiable assumptions about the joint distribution of the counterfactuals. However, posterior inferences on causal effects towards the tails of the birth weight distribution (e.g. among LBW infants) are not robust under the class of joint distributions that we considered. More specifically we found that, among LBW infants, the effect of micronutrient supplementation on birth weight is greatest and its estimate’s size is highly sensitive to  $k_\rho$ ; lower values of  $k_\rho$  correspond to a larger causal increase in birth weight.



**Fig. 5.** 95% posterior regions of  $F_0(W_i^{obs}, \theta_0^j)$  and  $F_1(W_i^{obs}, \theta_1^j)$  where  $F_0$  and  $F_1$  are the CDFs of the mixture of three normal distributions and  $\theta_0^j$  and  $\theta_1^j$  are the  $j$ th posterior sample of the vector of parameters of the mixture (—, corresponding empirical CDF): (a) control; (b) treatment

The posterior distributions of the population and causal 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 missing counterfactuals. In addition, we can easily explore the sensitivity of the posterior inferences to unverifiable assumptions about the joint distribution between the observed and the counterfactual variables.

To implement our approach we make several important assumptions. The first two (the stable unit treatment value assumption and random assignment) are justified by the randomization of the treatment assignment and the independence of the sampling units. Third we assume perfect compliance. The compliance for this trial was very good and did not depend on the treatment (Christian *et al.*, 2003a). The fourth and the fifth assumptions are in the logistic regression model for the probability of infant mortality as a function of the treatment indicator and the birth weight for the treatment received. Under the fourth assumption, we expect that the risk of mortality under the treatment would depend only on the actual birth weight and not on the birth weight for the intervention that is not received. This would be violated if there is a latent ‘growth potential’ for each child about which both  $W_i(0)$  and  $W_i(1)$  are informative and which is associated with mortality. This assumption is an avenue for further study. The fifth assumption, that the direct effect of the intervention on mortality is common to babies of all sizes, is consistent with the patterns in Fig. 1 but there is little statistical power to show otherwise. Finally the sixth and the seventh assumptions are about the associations between the observed and the missing counterfactuals and these associations cannot be estimated from the data. To deal with this unidentified problem we

- (a) use data on siblings to estimate lower bounds for  $k_\rho$  and  $\psi$  and use those as a guide for our prior choices and sensitivity analyses,
- (b) explore the sensitivity of estimated causal parameters with respect to choices for  $k_\rho$  and  $\psi$  and
- (c) compare inferences on causal parameters *versus* inferences on population parameters which are not affected by  $k_\rho$  and  $\psi$ .

The methodological development of this paper cuts across several contributions in quantile regression and counterfactual models. For example, we could have estimated the percentile-specific parameter  $\Delta_p^W$  by use of a quantile regression model of the form  $Q(p) = \alpha_p + \Delta_p^W Z_i$  (Koenker and Bassett, 1978) where  $Q(p)$  is the quantile function of  $W_i^{\text{obs}}$  and  $Z_i$  represents the treatment assignment. However, in this paper we extended the traditional definition of percentile-specific regression coefficients in two ways:

- (a) we introduce percentile-specific regression coefficients in the presence of post-treatment variables where the treatment effect on the dependent variable is allowed to vary with respect to the percentiles of an intermediate variable ( $\Delta_p^Y$ );
- (b) we introduce causal analogues of percentile-specific regression coefficients which vary with respect to the percentile of the counterfactual  $W_i(0)$  ( $\tau_p^W, \tau_p^Y$ ).

Estimation methods in quantile regression are based on finding the solution of a quantile regression minimization problem with a prespecified loss function (Koenker and Bassett, 1978). Bayesian analogues are described by Yu and Moyeed (2001). Our estimation approach for the percentile-specific parameters is simply based on transformations of the posterior samples of  $(W_i(0), W_i(1))$ . By modelling the marginal distributions of  $W_i(0)$  and  $W_i(1)$  as a mixture of normal distributions instead of a single normal distribution, we allow very flexible shapes for the percentile-specific treatment effects. The modest gain in goodness of fit, especially at the tails of the birth weight distribution, is shown in Fig. 2 and is supported by the posterior inferences on the variance components of the mixture model.

In the counterfactual model literature, Angrist *et al.* (1996) showed how instrumental variables can be embedded within the Rubin causal model for estimating an average causal effect in the presence of a binary post-treatment variable. They introduced five assumptions under which an instrumental variables estimator (Durbin, 1954) can be interpreted as the average

causal effect. The first two assumptions are the stable unit treatment value and the random assignment assumptions. The third assumption, which is called exclusion restriction, assumes that any effect of the treatment on the health outcome must be via an effect of the treatment on the post-treatment variable, i.e. there is no direct effect. We are not making this assumption: we use principal stratification to compare the different causal pathways on how the supplementation affects survival. In addition Angrist *et al.* (1996) assumed monotonicity in the post-treatment variable, i.e. that  $W_i(1) \geq W_i(0)$ ; we instead define a joint model for  $(W_i(0), W_i(1))$ .

By specifying a joint model for  $(W_i(0), W_i(1))$  which allows for the correlation  $\rho$  between the normalized percentiles of  $W_i(0)$  and  $W_i(1)$ , we provide a stochastic generalization of the rank preservation assumption (Efron and Feldman, 1991) which is similar to the assumption that was recently proposed by Dobbin and Louis (2003). More specifically, the hypothesis of rank preservation (also called percentiles invariance) implies that, for any group of participants, the birth weight percentiles would not be permuted if the group had been assigned to another treatment. In our model specification for the birth weights, the percentile invariance assumption leads to  $\rho_i = 1$  for all infants, which also implies that all the population parameters  $\Delta$  are equal to the causal parameters  $\tau$ .

The methodology that we described has broad applicability to a variety of situations in which one investigates a continuous post-treatment variable that potentially mediates a binary response of interest. For example, similar issues arise in cancer trials that evaluate both tumour growth and survival. In these studies it is likely that there are both mediated and direct effects of treatments, and that these effects may vary across the distribution of tumour growths. 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.

The estimation of treatment effects by percentile of the birth weight distribution has public health significance. In the case-study that was presented here, the treatment increased the birth weight of smaller babies and had no apparent effect on larger babies. If it had increased the size of the larger infants, both the infants and their mothers might have been at higher risk of mortality given the absence of obstetrical care in rural communities. In such a situation, it would be necessary to predict those mothers who are likely to have larger infants and to exclude them from intervention programmes. However, although maternal prepregnancy nutritional status, weight gain during pregnancy and other factors are strong determinants of LBW, their ability to predict infants who are likely to be born with LBW is still uncertain.

Currently recommendations exist for supplementing women with iron-folic acid during pregnancy in developing countries. The Nepal study (Christian *et al.*, 2003a) demonstrated that, beyond reducing anaemia, 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 differential beneficial effects of an intervention depending on where in the distribution the programme participants fall and that an overall effect size may

- (a) underestimate the maximum likely benefit in the most malnourished individuals and
- (b) incorrectly assume benefits where none exist and potentially mask harm in the better nourished individuals.

## Acknowledgements

Funding for Francesca Dominici and Scott L. Zeger was provided by a grant from the National Institute for Environmental Health Sciences (ES012054-01) and by the National Institute for Environmental Health Sciences Center in Urban Environmental Health (P30 ES 03819). We thank Dr Constantine Frangakis and Dr Thomas Louis for comments.

## References

- Albert, J. H. and Chib, S. (1993) Bayesian analysis of binary and polychotomous response data. *J. Am. Statist. Ass.*, **88**, 669–679.
- Angrist, J. D., Imbens, G. W. and Rubin, D. B. (1996) Identification of causal effects using instrumental variables (with discussion). *J. Am. Statist. Ass.*, **91**, 444–472.
- Begg, C. B. and Leung, D. H. Y. (2000) On the use of surrogate end points in randomized trials. *J. R. Statist. Soc. A*, **163**, 15–24.
- Buyse, M. and Molenberghs, G. (1998) Criteria for the validation of surrogate endpoints in randomized experiments. *Biometrics*, **54**, 1014–1029.
- Caulfield, L., Zavaleta, N., Figueroa, A. and Leon, Z. (1999) Maternal zinc supplementation does not affect size at birth and pregnancy duration in Peru. *J. Nutr.*, **129**, 1563–1568.
- Ceesay, S., Prentice, A., Cole, T., Foord, F., Weaver, L. and Poskitt, E. (1997) Effects on birth weight and perinatal mortality of maternal dietary supplements in rural Gambia: 5 randomized controlled trials. *Br. Med. J.*, **315**, 786–790.
- Chib, S. and Greenberg, E. (1998) Analysis of multivariate probit models. *Biometrika*, **85**, 347–361.
- Child Health Research Project (1996) Special Report, Reducing perinatal and neonatal mortality, Report of a meeting. *Child Health Research Project*.
- 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. *Br. Med. J.*, **326**, 1–6.
- Christian, P., West, K., Khatry, S., Leclercq, 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. *Am. J. Clin. Nutr.*, **78**, 1194–1202.
- Cole, S. and Hernan, M. (2002) Fallibility in estimating direct effects. *Int. J. Epidemiol.*, **31**, 163–165.
- Cowles, M. K. (2002) Bayesian estimation of the proportion of treatment effect captured by a surrogate marker. *Statist. Med.*, **21**, 811–834.
- Cox, D. (1984) Interaction. *Int. Statist. Rev.*, **52**, 1–31.
- Daniels, M. J. and Hughes, M. D. (1997) Meta-analysis for the evaluation of potential surrogate markers. *Statist. Med.*, **16**, 1965–1982.
- Diebolt, J. and Robert, C. P. (1994) Estimation of finite mixture distributions through Bayesian sampling. *J. R. Statist. Soc. B*, **56**, 363–375.
- Dobbin, K. K. and Louis, T. A. (2003) Accommodating stochastic departures from percentiles invariance in causal models. *J. R. Statist. Soc. B*, **65**, 837–849.
- Doksum, K. A. and Sievers, G. L. (1976) Plotting with confidence: graphical comparisons of two populations. *Biometrika*, **63**, 421–434.
- Dominici, F., Cope, L., Naiman, D. and Zeger, S. L. (2005a) Smooth quantile ratio estimation (SQUARE). *Biometrika*, **92**, 543–557.
- Dominici, F., Zeger, S. L., Parmigiani, G., Katz, J. and Christian, P. (2005b) Does the effect of micronutrient supplementation on neonatal survival vary with respect to the percentiles of the birth weight distribution? *Technical Report*. Department of Biostatistics, Johns Hopkins University, Baltimore.
- Durbin, J. (1954) Errors in variables. *Rev. Int. Statist. Inst.*, **22**, 23–32.
- Efron, B. and Feldman, D. (1991) Compliance as an explanatory variable in clinical trials. *J. Am. Statist. Ass.*, **86**, 9–17.
- Fraley, C. and Raftery, A. E. (2002) Model-based clustering, discriminant analysis, and density estimation. *J. Am. Statist. Ass.*, **97**, 611–631.
- Frangakis, C. E. and Rubin, D. B. (2002) Principal stratification in causal inference. *Biometrics*, **58**, 21–29.
- Freedman, L., Graubard, B. and Schatzkin, A. (1992) Statistical validation of intermediate endpoints for chronic diseases. *Statist. Med.*, **11**, 167–179.
- 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.
- Gelman, A., Carlin, J., Stern, H. and Rubin, D. (1995) *Bayesian Data Analysis*. London: Chapman and Hall.
- Holland, P. (1986) Statistics and causal inference. *J. Am. Statist. Ass.*, **81**, 945–960.
- Kadane, J. (1975) The role of identification in Bayesian theory. In *L. J. Savage Memorial Volume: Studies in Bayesian Statistics and Econometrics* (eds S. Fienberg and A. Zellner), pp. 175–191. Amsterdam: North-Holland.

- Katz, J., Christian, P., Dominici, F. and Zeger, S. L. (2005) Treatment effects of maternal micronutrient supplementation vary by percentiles of the birth weight distribution in rural Nepal. *Technical Report*. Department of Biostatistics, Johns Hopkins University, Baltimore.
- 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. *Int. J. Epidemiol.*, **30**, 802–807.
- 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. *Am. J. Clin. Nutr.*, **71**, 1570–1576.
- 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. *Am. J. Clin. Nutr.*, **71**, 1570–1576.
- Koenker, R. and Bassett, G. J. (1978) Regression quantiles. *Econometrica*, **46**, 33–50.
- Koenker, R. and Geling, O. (2001) Reappraising medfly longevity: a quantile regression survival analysis. *J. Am. Statist. Ass.*, **96**, 458–468.
- van der Laan, M. and Petersen, M. (2004) Estimation of direct and indirect causal effects in longitudinal studies. *Technical Report 1155*. Division of Biostatistics, University of California, Berkeley.
- Lechtig, A., Yarbrough, C., Delgado, H., Habicht, J., Marorelli, R. and Klein, R. (1975) Influence of maternal nutrition on birth weight. *Am. J. Clin. Nutr.*, **28**, 1223–1233.
- Leung, D. H.-Y. (2001) Statistical methods for clinical studies in the presence of surrogate end points. *J. R. Statist. Soc. A*, **164**, 485–503.
- Liang, K.-Y., Zeger, S. L. and Qaqish, B. (1992) Multivariate regression analyses for categorical data (with discussion). *J. R. Statist. Soc. B*, **54**, 3–40.
- McIntire, D., Bloom, S., Casey, B. and Leveno, K. (2001) Birth weight in relation to morbidity and mortality among newborn infants. *New Engl. J. Med.*, **340**, 1234–1238.
- Molenberghs, G., Geys, H. and Buyse, M. (2001) Evaluation of surrogate endpoints in randomized experiments with mixed discrete and continuous outcomes. *Statist. Med.*, **20**, 3023–3038.
- Pearl, J. (2000) *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press.
- Prentice, R. L. (1986) Binary regression using an extended beta-binomial distribution, with discussion of correlation induced by covariate measurement errors. *J. Am. Statist. Ass.*, **81**, 321–327.
- 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. *Br. Med. J.*, **327**, 254–257.
- Rasmussen, K. M. (2001) Is there a causal relationship between iron deficiency anemia and weight at birth, length of gestation and perinatal mortality? *J. Nutr.*, **131**, 590S–603S.
- Robins, J. (1989) The control for confounding by intermediate variables. *Statist. Med.*, **8**, 679–701.
- Robins, J. M. and Greenland, S. (1992) Identifiability and exchangeability for direct and indirect effects. *Epidemiology*, **3**, 143–155.
- Rubin, D. B. (1978) Bayesian inference for causal effects: the role of randomization. *Ann. Statist.*, **6**, 34–58.
- Tanner, M. A. (1991) Tools for statistical inference—observed data and data augmentation methods. *Lect. Notes Statist.*, **67**.
- Tanner, M. A. and Wong, W. H. (1987) The calculation of posterior distributions by data augmentation. *J. Am. Statist. Ass.*, **82**, 528–550.
- Tierney, L. (1994) Markov chains for exploring posterior distributions (with discussion). *Ann. Statist.*, **22**, 1701–1762.
- West, J. J., Katz, J., Khatry, S. K., LeClerq, S. 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. *Br. Med. J.*, **318**, 570–575.
- Xu, J. and Zeger, S. L. (2001) The evaluation of multiple surrogate endpoints. *Biometrics*, **57**, 81–87.
- Yu, K. and Moyeed, R. A. (2001) Bayesian quantile regression. *Statist. Probab. Lett.*, **54**, 437–447.