Original Contribution

Birth Weight and Mortality: Causality or Confounding?

Olga Basso¹, Allen J. Wilcox¹, and Clarice R. Weinberg²

¹ Epidemiology Branch, National Institute of Environmental Health Sciences, National Institutes of Health, US Department of Health and Human Services, Research Triangle Park, NC.
² Biostatistics Branch, National Institute of Environmental Health Sciences, National Institutes of Health, US Department of Health and Human Services, Research Triangle Park, NC.

Received for publication September 1, 2005; accepted for publication December 14, 2005.

The association between birth weight and mortality is among the strongest seen in epidemiology. While preterm delivery causes both small babies and high mortality, it does not explain this association. Fetal growth restriction has also been proposed, although its features are unclear because it lacks a definition independent of weight. If, as some postulate, birth weight is not itself on the causal path to mortality, its relation with mortality would have to be explained by confounding factors that decrease birth weight and increase mortality. In this paper, the authors explore the characteristics such confounders would require in order to achieve the observed association between birth weight and mortality. Through a simple simulation, they found that the observed steep gradient of risk for small babies at term can be produced by a rare condition or conditions (with a total prevalence of 0.5%) having profound effects on both fetal growth (–1.7 standard deviations) and mortality (relative risk = 160). Candidate conditions might include malformations, fetal or placental aneuploidy, infections, or imprinting disorders. If such rare factors underlie the association of birth weight with mortality, it would have broad implications for the study of fetal growth restriction and birth weight, and for the prevention of infant mortality.

birth weight; confounding factors (epidemiology); fetal growth retardation; infant, small for gestational age; mortality

Editor’s note: An invited commentary on this article appears on page 312, and the authors’ response is published on page 315.

The association between birth weight and infant survival is among the strongest seen in the whole of epidemiology. Babies weighing less than 1,500 g have a mortality risk at least 100-fold higher than babies at the optimum weight (the weight associated with the lowest mortality). While this association is highly robust, its underlying biologic mechanisms are not clearly understood. If birth weight itself were the cause of death, the association would be self-explanatory. If, on the other hand, birth weight were not on the causal path to mortality, as some authors suggest (1, 2), confounding would be the underlying mechanism. Both mechanisms may coexist. In this paper, we explore what characteristics would be required of a confounding factor if birth weight were not on the path to mortality.

ASSOCIATION OF BIRTH WEIGHT WITH MORTALITY

The data

We have restricted the discussion to neonatal mortality (death within 28 days of birth), although similar patterns are found with infant mortality. To illustrate our examples, we used the 1995–2000 linked data from vital statistics for the United States (3). We excluded babies born of non-US residents (0.1 percent), multiple births (2.9 percent), and babies for whom birth weight was missing (0.1 percent). Approximately 3 percent of all neonatal deaths could not be linked to a birth and were ignored. To estimate cohort mortality, rather than period mortality, we excluded births occurring in
December 2000 and deaths in 1995 of babies born in 1994. In 1 percent of the records, gestational age was missing.

Figure 1 shows birth-weight-specific mortality for the United States in 1960 and 2000. Note that the scale of mortality is logarithmic: birth-weight-specific mortality rises exponentially throughout the lower birth-weight range. A weaker increase in mortality is seen among the heaviest babies. This reverse-J pattern of weight-specific mortality is apparently universal, being found in diverse populations (4–9). Between 1960 and 2000, neonatal mortality in the United States declined more than 75 percent (from 153 to 37 per 10,000 births). This decline is expressed as a relatively uniform decrease in mortality (on a log scale) at each birth weight (figure 1). Despite the decrease in absolute mortality, the steep gradient of mortality among small babies has persisted over time.

The role of preterm delivery in weight-specific mortality

Figure 1 data comprise babies of all gestational ages. Since preterm delivery is associated with both small size and high mortality, much of the pattern shown in figure 1 could, in principle, reflect the effects of preterm delivery. Figure 2 shows US weight-specific mortality (1995–2000) for all babies and also for babies born at 40 weeks (defined by the date of the last menstrual period). Although neonatal mortality is reduced to 8 per 10,000 at 40 weeks (compared with 38 per 10,000 overall), the restriction to babies born at 40 weeks does not reduce the slope of the curve. Indeed, the steep gradient has been shown to persist within every stratum of gestational age (10).

Errors in gestational age could artificially increase the slope by preferentially including preterm babies among the smallest ones. To explore the role of error, we further restricted our analysis to babies with a last-menstrual-period estimate of 40 weeks of gestation (solid line) by last menstrual period. Data for babies weighing <1,000 g were grouped in one category (no babies weighed <750 g at 40 weeks). Excluded were 196 babies weighing ≥6,000 g, and those weighing 5,250–5,999 g were grouped in a single category. A total of 5,049,104 babies remained. Values on the x-axis represent the midvalue for the category.

The role of fetal growth restriction in weight-specific mortality

Fetal growth restriction is another possible explanation for high mortality at low birth weights. Clinical attention
to fetal growth restriction has focused on the smallest babies, which are the ones most likely to be growth restricted. This emphasis on the smallest babies is expressed in definitions of “fetal growth restriction” as the lowest 10th or 5th percentile of birth weight at a given gestational age. There is, however, no agreed-upon clinical definition of growth restriction other than size itself.

An alternative way to conceptualize “fetal growth restriction” is as a failure of the fetus to achieve its target birth weight (11, 12). Such babies would not necessarily be among the smallest 10 percent. The size of these growth-restricted babies, while smaller than biologically intended, could still span a wide spectrum of weights.

For the remainder of this discussion, we use “fetal growth restriction” not to designate a category of small babies below some percentile but to describe any interference with fetal growth regardless of the baby’s absolute weight at birth. A factor that causes fetal growth restriction affects babies regardless of their biologically intended weight (i.e., there is a shift of the whole distribution of birth weights). As a result, mean weight decreases and the proportion of small babies increases.

A factor that restricts fetal growth and also increases mortality would act as a confounder to produce an association between birth weight and mortality. What characteristics would be required of such a factor, if we assume that birth weight itself has no direct effect on mortality?

ESTIMATING THE EFFECT OF FETAL GROWTH RESTRICTION ON MORTALITY

Maternal smoking

Cigarette smoking is an example of a factor that causes fetal growth restriction and increases infant mortality (1). Babies born at term to smokers have the familiar Gaussian distribution of birth weights, but the distribution is shifted downward by 200 g (about 0.4 standard deviations), with no change in the standard deviation itself. On the basis of this shifted weight distribution, it could be stated that all babies born to smokers are growth restricted, even though their birth weights cover nearly the full range of possible weights. In our data, maternal smoking also increases neonatal mortality by about 30 percent.

Let us assume a population of term babies in which 20 percent of the mothers are smokers. We also assume that birth weight itself, prior to any interference, is unrelated to mortality and that no other factors are at work. This setting allows us to determine how much a factor such as smoking can contribute to the gradient of weight-specific mortality.

Figure 3 (left) shows the resulting hypothetical birthweight distributions for babies of smokers and nonsmokers (with the babies of smokers shifted to lower birth weights) and with a higher mortality among the babies of smokers. Figure 3 (right) combines the two groups of babies and shows the resulting pattern of weight-specific mortality. While smoking produces a weak gradient of mortality, it does not come close to approaching the observed gradient in real populations (e.g., figure 1). Other factors with similar and independent effects on birth weight and mortality (e.g., social class) would presumably produce a similarly shallow curve. Even by summing many such factors, it would be virtually impossible to achieve the gradient observed for real populations.

Factor X1

A confounding factor that explains the observed association of birth weight with mortality must have properties quite different from smoking or other common causes of fetal growth restriction. In this paper, we use simulations to estimate those properties. For the sake of simplicity, we hypothesize a scenario in which all babies are born at term (i.e., in the 40th week of gestation), so as to remove preterm delivery as a consideration. Our assumptions follow.

1. We assume a “natural” Gaussian distribution of birth weights, determined largely by the biologic potential of each fetus (target birth weight). We express this Gaussian distribution in standardized units, with a mean of 0 and a standard deviation of 1. Target birth weight and mortality are unrelated in the absence of any interference with the fetus’ growth (figure 4, left). We set the baseline mortality at 0.4 per 1,000 for this example, close to the observed mortality at the lowest point on the empirical mortality curve (“optimum birth weight”).

2. We assume a condition (X1) that decreases birth weight and increases mortality enough to produce the observed gradient of mortality for small babies. After experimenting with a range of values, we find that these effects must be surprisingly extreme. In the example presented here, X1 decreases birth weight by −1.7 standard deviations (about 770 g) and increases mortality to 160 times the baseline, with prevalence of 0.5 percent (figure 4, center).

3. We assume a second condition (X2) that increases mortality and also increases birth weight, to produce the hook of the J shape. We find a good fit to the data under the following conditions: X2 has a prevalence of 0.3 percent, X2 changes birth weight by the same amount as X1 but in the opposite direction (1.7 standard deviations), and X2 increases mortality 16-fold (figure 4, center).

4. In this simple example, we assume that, for babies affected by X1 or X2, mortality remains independent of birth weight (i.e., the mortality curve is flat).

The right panel of figure 4 shows the weight-specific mortality curve that results from these assumptions. Figure 5 compares the simulated mortality curve with that for US babies born at 40 weeks of gestation (with standard deviation estimated by identifying the predominant distribution (13)). (The formula used for the simulation is provided in appendix 1.) Thus, the observed pattern of weight-specific mortality can be replicated with very simple assumptions. While other combinations of parameters produce the observed gradient of risk with low weights at 40 weeks, all seem to require similarly extreme values. That is, to achieve the steep gradient in mortality, X1 must strongly restrict fetal growth and confer high mortality, while being relatively rare. In the following section, we discuss the inflexibility of these requirements for X1.
Sensitivity analysis

In figure 6, we explore the sensitivity of this model to changes in the parameters of $X_1$. The four panels show the original simulated curve (solid black) with an alternate in which one of the four parameters has been changed. Doubling the fraction of babies with $X_1$ shifts the mortality curve to the left (figure 6, top left); reducing the fetal-growth-restricting effect of $X_1$ (from $\frac{-1.7}{\text{standard deviation}}$ to $\frac{-1.0}{\text{standard deviation}}$) leads to a shallower slope (figure 6, top right). Decreasing the impact of $X_1$ on mortality changes both the shape and the slope of the curve (figure 6, bottom left). Finally, changing the baseline mortality causes an overall shift in the mortality curve upward or downward (figure 6, bottom right)—strikingly similar to the change observed over time in the US data (figure 1). In sum, we find that relatively minor deviations from the values assumed in our simulation produce distinct deviations from the empirical data.

DISCUSSION

Although our scenario is clearly oversimplified, the model is able to replicate the observed mortality curve. Furthermore, our simulations suggest that the requirements for the parameters of $X_1$ are extreme and fairly narrow. Specifically, the prevalence of $X_1$ at term could be only a few per 1,000 births under the hypothesized scenario. Such an extremely low prevalence for $X_1$ (and $X_2$) is consistent with the empirical distribution of birth weight. In virtually all populations, the distribution of birth weights at 40 weeks is close to a pure Gaussian curve. Shifting 0.5 percent of the babies toward the lower tail and 0.3 percent toward the upper tail of a Gaussian distribution would have a virtually imperceptible effect on this shape.

The extremely high relative risk associated with $X_1$ may appear unrealistic, but it is not so large in absolute terms: at 40 weeks, the relative risk of 160 is associated with an absolute mortality of 6.4 percent. This level of mortality at low birth weights can be achieved only with a high relative risk, if only one factor is postulated. We explored whether several weaker factors might be involved rather than one strong one. If these weaker factors were highly correlated, it would be conceptually equivalent to our proposed scenario. If, on the other hand, the factors in question were independent of one another, a larger number of babies would have to be affected in order to obtain a small fraction having all the factors necessary to confer the high risk. In our simulations, we were not able to come close to reproducing the mortality curve at term with a combination...
of three independent factors weaker than $X_1$. Furthermore, the higher prevalence required for each of these weaker factors led to profound distortions in the Gaussian distribution of birth weight.

**Assumptions of the model**

We have explored the characteristics that a confounder would be required to have in order to produce the observed association between birth weight and mortality. We believe it is not unreasonable to assume that the biologic potential for fetal growth could be unrelated to mortality. One argument that might be raised against this assumption is that there is some absolute lower limit of viable birth weight, and babies who are constitutionally small may be especially susceptible to risk of death if they are pressed closer to the limits of viability by fetal growth restriction. While this is possible, the empirical evidence suggests that some factors can shift the entire distribution of birth weight toward lower values without any apparent effect on mortality, as appears to be the case with babies born at high altitude ($1$).

The assumption of independence between birth weight and adverse outcome may not hold when long-term health, rather than death, is the outcome. Additionally, while both factors specified in this paper may have long-term health consequences, our estimated frequencies are too low to explain the empirical relation observed between birth weight and certain diseases later in life ($14$). However, weaker and more prevalent confounding factors may be related to later outcomes with a mechanism similar to that described here.

Our simulations also assume that the effects of $X_1$ are uniform, that is, for all affected babies, the shift in birth weight is identical and the relative risk is identical. This assumption is unlikely to be true. We can refine the model by substituting independent distributions for the constant effects, with no essential modification of the model and no major changes in the results. (The same applies to $X_2$.)

One could question the assumption that babies within categories of $X_1$ and $X_2$ have a constant mortality independent of their birth weight (or independent of the degree of fetal growth restriction, if a distribution of shifts is used). While it is possible to assume that mortality is proportional to the shift, it is not necessary. Our simulation shows that a simple model can generate the weight-specific mortality curve without any effect of birth weight on mortality. Thus, under the simple assumption of $X_1$, the observed strong relation between low birth weight and mortality could be due entirely to confounding.

Note that, in our simulation, as in reality, the minimum mortality occurs at a weight slightly above the mean. Our model achieves this by assigning a greater impact to the factor decreasing birth weight than to the factor increasing fetal growth restriction, if a distribution of shifts is used. While it is possible to assume that mortality is proportional to the shift, it is not necessary. Our simulation shows that a simple model can generate the weight-specific mortality curve without any effect of birth weight on mortality. Thus, under the simple assumption of $X_1$, the observed strong relation between low birth weight and mortality could be due entirely to confounding.

**Applicability of the model at earlier gestational ages**

Simulations applied to earlier gestational ages are complicated by the increased influence of errors in gestational age.
Such errors are evident in the data on last menstrual period and also when using the "clinical estimate" of gestation. We explored these problems by applying external estimates of gestational-age-specific birth-weight means and standard deviations (15) to calculate a corrected birth-weight distribution that excluded values lower than \( \mu \pm 5 \) standard deviations and higher than 4 standard deviations. We applied our simulation (using the odds ratio to accommodate the higher baseline mortality; refer to appendix 1) to several categories of preterm gestation and were able to achieve a reasonable fit for most (refer to appendix 2 and appendix figure 1 for an example). In general, the estimated prevalence of \( X_1 \) and \( X_2 \) was higher among the preterm births. However, the combination of errors in gestational age, the decreasing sample size at earlier gestational ages, and the use of external parameters for birth weight make us less certain about the estimated parameters of \( X_1 \) and \( X_2 \) in preterm strata because of the sensitivity of these estimates to the standard deviation and to the slope of the mortality curve.

**Biologic mechanisms**

For fetal growth restriction to account for the steep gradient of mortality with low birth weights, it must be of a form far more severe (and rare) than the factors usually discussed (16). Candidates for this kind of fetal growth restriction would be rare conditions such as certain congenital malformations, aneuploidy, some fetal or maternal infections (17–19), severe maternal kidney disease (17), confined placental mosaicism (12), uniparental disomy (12), or disorders of imprinting (20, 21). Kingdom et al. (22) describe a rare form of fetal growth restriction that likely originates in the formation of the placental villous tree, although this condition would likely not be compatible with survival to term. \( X_1 \) might constitute a constellation of rare mechanisms whose prevalences change throughout gestation because of differential attrition. As-yet unrecognized mechanisms may also be part of \( X_1 \).

Candidates for \( X_2 \) may include overexpression of insulin-like growth factor 2 (23) or untreated gestational diabetes (24). Size itself may contribute to the increased mortality of large babies by increasing the risk of injury at birth, although the mortality pattern at early gestations (when babies are smaller in absolute terms) suggests that this cannot be the sole mechanism. We are, however, less concerned about \( X_2 \) because its contribution to total mortality is limited. In our simulation, about 4 percent of all deaths at term were due to \( X_2 \), compared with about 44 percent due to \( X_1 \).

**Implications**

If the steep gradient of the mortality curve is caused by a small number of babies at very high risk, one implication is that, among babies at the lowest 10 percent of weight for gestation, the great majority would be healthy without \( X_1 \). In our simulation, only 3 percent of babies born at 40 weeks at a birth weight below the 10th percentile would have \( X_1 \). Even if the 1st percentile were selected as the criterion for at-risk babies, only 13 percent would be affected by \( X_1 \). In a search for pathologic conditions associated with \( X_1 \), care would have to be taken to maximize the likelihood of \( X_1 \) in the sample, for example, by restricting to the smallest 2 or 3 percent of babies, or to small babies among the preterm (where the prevalence of \( X_1 \) may be larger), or to perinatal deaths (which presumably include a high proportion of \( X_1 \) babies).

If we pursue the speculation that \( X_1 \) explains the high gradient of risk for small babies, it would follow that birth weight itself may not be a suitable target for direct intervention. Focus would shift to factors that affect baseline mortality risk. While it would be worthwhile to identify the factors that comprise \( X_1 \), it is also important to recognize that the great historical declines in neonatal mortality have apparently occurred solely through decreases in baseline mortality (figure 1). Thus, factors that affect baseline mortality may generally be more amenable to modification and prevention than \( X_1 \).

In sum, we propose that the high mortality of small babies can, at least in theory, be completely explained by the presence of a persistent, rare, and unmeasured confounder.

*Am J Epidemiol* 2006;164:303–311
FIGURE 6. Sensitivity analyses for parameters of the model. The solid black curve represents the one fitted with the parameters described in the text. The dotted curve differs by one parameter in each panel. Top left: prevalence of $X_1$ is doubled from 0.5% to 1%. Top right: shift in birth weight due to $X_1$ is reduced from $1.6$ standard deviations to $1$ standard deviation. Bottom left: effect of $X_1$ on mortality is reduced from a relative risk of 160 to a relative risk of 100. Bottom right: baseline mortality is quadrupled, from 4 per 10,000 to 16 per 10,000. Refer to the text for a description of $X_1$. 

Am J Epidemiol 2006;164:303–311
ACKNOWLEDGMENTS

This research was supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences.

The authors are indebted to colleagues who challenged and encouraged them during the development of these ideas. Drs. Donna Baird, David Haig, Michael Kramer, Jørn Olsen, David Savitz, and Rolf Skjærven reviewed earlier versions of the manuscript, and their comments helped improve the paper substantially. Drs. Glinda Cooper, David Dunson, Beth Gladen, Freya Kamel, Stephanie London, Matthew Longnecker, and David Umbach offered important critical comments.

Conflict of interest: none declared.

REFERENCES

2. Haig D. Meditations on birth weight: is it better to reduce the variance or increase the mean? Epidemiology 2003;14:490–2. (Erratum: Epidemiology 2003;14:6:32).

APPENDIX 1

Mathematical Formulation of the Simulation

\[ f(z) = \frac{\pi_{11} \exp \left( -\frac{(z - \alpha)^2}{2} \right) R_{11} + \pi_{01} \exp \left( -\frac{(z - \beta)^2}{2} \right) R_{01}}{\pi_{11} \exp \left( -\frac{(z - \alpha)^2}{2} \right) + \pi_{01} \exp \left( -\frac{(z - \beta)^2}{2} \right) + \pi_{10} \exp \left( -\frac{(z - \gamma)^2}{2} \right) + \pi_{00} \exp \left( -\frac{(z - \delta)^2}{2} \right)} \]

where

\( f(z) = \) risk of death at value \( z \) (standardized birth weight, with \( \mu = 0 \) and \( \sigma = 1 \))

\( i \) = presence (1) or absence (0) of \( X_1 \)

\( j \) = presence (1) or absence (0) of \( X_2 \)

\( \pi_{ij} \) = frequency of babies with a given combination of \( X_1 \) and \( X_2 \)

\( R_{ij} \) = risk of death for babies with a given combination of \( X_1 \) and \( X_2 \)

\( z \) = value of \( z \) (standardized unit) for which the risk is calculated

\( \alpha, \beta, \gamma, \delta \) = shift in birth weight (standardized units) due to \( X_1, X_2 \)
Note that, when the relative risk is used as the relative measure of risk, the risk of death within subgroups may exceed 1. At 40 weeks of gestation, we constrained the risk to be 1 (instead of 1.02) in the subgroup having both $X_1$ and $X_2$. The model at 35 weeks was generated by using the odds ratio instead of the relative risk, which changes only how $R_{ij}$ is calculated.

With relative risk ($RR$), $R_{ij} = \text{baseline} \times RR_{ij}$

With odds ratio ($OR$), $R_{ij} = \frac{(\text{baseline} \times OR_{ij})}{\left[ (1 - \text{baseline}) + (\text{baseline} \times OR_{ij}) \right]}$,

where $RR_{ij}$ and $OR_{ij}$ represent the relative measure of risk calculated for the whole subgroup (depending on whether $X_1$, $X_2$, both, or neither are present).

**APPENDIX 2**

**An Example using Gestational Week 35**

We used the clinical estimate of gestational age, after excluding 10 babies whose birth weight was <500 g and 959 whose birth weight was >4,250 g; thus, 292,487 babies remained. We used the mean and standard deviation as proposed by Kramer et al. (15), after calculating a pooled value for boys and girls (throughout the analyses, birth weights were reported in 250-g categories because of the large quantity of data). We used the odds ratio to model the risk considering the higher baseline mortality at 35 weeks. Appendix figure 1 shows the fit of the model, with the frequency of $X_1$ set at 1.3 percent (with a shift of $-1.40$ standard deviations and an odds ratio of 110) and the frequency of $X_2$ set at 0.4 percent (with a shift of $1.35$ standard deviations and an odds ratio of 17).