© International Epidemiological Association 2001 Printed in Great Britain *International Journal of Epidemiology* 2001;**30**:1233–1241

REFORMULATIONS

On the importance—and the unimportance—of birthweight

Allen J Wilcox

Birthweight is one of the most accessible and most misunderstood variables in epidemiology. A baby's weight at birth is strongly associated with mortality risk during the first year and, to a lesser degree, with developmental problems in childhood and the risk of various diseases in adulthood. Epidemiological analyses often regard birthweight as on the causal pathway to these health outcomes. Under this assumption of causality, birthweight is used to explain variations in infant mortality and later morbidity, and is also used as an intermediate health endpoint in itself. Evidence presented here suggests the link between birthweight and health outcomes may not be causal. Methods of analysis that assume causality are unreliable at best, and biased at worst. The category of 'low birthweight' in particular is uninformative and seldom justified. The main utility of the birthweight distribution is to provide an estimate of the proportion of small preterm births in a population (although even this requires special analytical methods). While the ordinary approaches to birthweight are not well grounded, the links between birthweight and a range of health outcomes may nonetheless reflect the workings of biological mechanisms with implications for human health.

Keywords Birthweight, fetal growth, gestational age, low birthweight, intrauterine growth retardation, small for gestational age, infant mortality, analytical methods

Accepted 30 July 2001

There are thousands of research papers on birthweight, with hundreds more appearing every year. Why is birthweight such a popular topic? One factor is that birthweight data are precisely recorded, free (through vital statistics), and available in vast numbers. A second factor is that birthweight is an extremely powerful predictor of an individual baby's survival. In general, the lower the weight, the higher a baby's risk of infant mortality.[1] A third factor is that, on a population level, mean birthweight is associated with infant mortality. Groups with lower mean birthweight often have higher infant mortality (e.g. the infants of mothers who smoke, or of mothers with lower socioeconomic status).[2,3] Finally, birthweight is associated with health outcomes later in life. Asthma, low IQ, and hypertension have all been reported to be more common among those who were small at birth.[4–6]

The strong association of birthweight with infant mortality is the usual focus of birthweight research, with the assumption that birthweight is a major determinant of infant survival. Researchers track population trends in birthweight, assuming these to have implications for infant mortality trends.[7] In the US, interventions to increase birthweight are recommended as a strategy to improve infant mortality.[8–10]

This commentary questions the causal role of birthweight in its association with health outcomes. The structure of this paper is as follows. The first section reviews the history of research on birthweight as a cause of infant mortality. The second describes the basic features of birthweight and its association with infant mortality. The third section proposes an alternative hypothesis of a non-causal association between birthweight and health outcomes. Under this alternative, birthweight is of limited importance as an end-point in itself, and is inconsequential in the analysis of infant mortality or other outcomes. The fourth and final section discusses implications for epidemiological research.

A Short History of Birthweight

For most of the previous century, birthweight has been treated as a dichotomy. 'Low birthweight' (LBW) is the category of babies weighing less than 2500 g at birth, and 'normal birthweight' is all the rest. For many years, the presumed reason for babies to be born LBW was their preterm delivery. Indeed, the terms 'LBW' and 'premature' were used interchangeably in the scientific literature from the 1920s to the 1960s.

Epidemiology Branch, National Institute of Environmental Health Sciences, Durham NC 27709, USA. E-mail: WILCOX@NIEHS.NIH.GOV

However, not all small babies are premature, and not all premature babies are small. An accumulation of epidemiological data during the 1950s and 1960s finally made this distinction clear. In 1961, the World Health Organization recommended that LBW no longer be used as the official definition of prematurity.[11] By the 1970s, most researchers were complying, although as late as 1977 a book on LBW was titled *The Epidemiology of Prematurity*.[12] Perinatal epidemiologists now avoid the word 'premature' altogether, preferring the label 'preterm' for a baby born too early.

As researchers began to recognize that LBW and preterm are not synonymous, they faced an uncomfortable new problem. Term babies born at less than 2500 g nonetheless have a high risk of mortality. What accounts for this risk, if not preterm delivery?

This gap was filled by the invention of a new disease—intrauterine growth retardation (IUGR). The usual definition of IUGR is 'small for gestational age' (SGA), the lightest 10% in each gestational age stratum. Under the percentile definition, the vast majority of IUGR babies are born at term. (This is simply a function of definition: under a percentile formula, the category of IUGR contains the same small per cent of preterm births as is present in the general population.) Taken as a whole, IUGR babies correspond closely with the set of LBW babies at term, and provides these LBW babies with a 'diagnosis'. Thus, the creation of an entity called IUGR effectively preserved LBW as a group of babies with 'preventable' ailments. Small babies who are not preterm are 'growth retarded'.

This convenient solution to the problem of term LBW infants led to the rapid acceptance of the concept of IUGR during the 1970s. According to PubMed, the number of papers about IUGR swelled between 1970 and 1979 from a handful to more than 200 a year. In fact, this was not a new research area but a shift within LBW research from one label ('prematurity') to two ('preterm' and 'IUGR').

Popular assumptions about LBW

The dichotomization of birthweight is deeply entrenched in public health research. Why have researchers been so determined to cling to this strategy? This practice rests on several assumptions about LBW.

LBW causes infant mortality

In the first year of life, LBW babies are typically 20 or more times more likely to die than heavier babies.[13] The sheer strength of this association with mortality is regarded as evidence of its causality.

The per cent LBW in a population is an indicator of infant risk

Infant death is rare (at least in developed countries), so researchers need a more prevalent surrogate indicator of perinatal risk. Low birthweight serves this purpose nicely. Furthermore, under this assumption, the causes of LBW themselves become topics of investigation.

LBW is preventable

If LBW is caused by either preterm delivery or fetal growth retardation, then LBW is presumably completely preventable. Thus, LBW provides a target for interventions to improve infant survival. The prevention of LBW is an explicit part of US public health policy to decrease infant mortality.[8]

While these assumptions about LBW are generally accepted, not all aspects of LBW neatly fit into them. For example, groups with a larger per cent of LBW babies do not invariably have the greater risk. A well-known example is the comparison of female and male babies.[14] But the most telling contradiction is the 'low birthweight paradox'.

The LBW paradox

Populations with a higher per cent of LBW often have higher rates of infant mortality. This supports the notion that LBW is a useful surrogate of population risk. However, there is an odd thing about LBW babies in high-risk populations—they usually have lower mortality than LBW babies in better-off populations. This is the LBW paradox, and its history is entwined with one of the most famous controversies in the history of epidemiology: the debate over the causal role of cigarette smoking.

In the 1950s, researchers found that mothers who smoked had smaller babies. By the 1960s, there was evidence that babies of these mothers also had higher infant mortality. But the effect of mother's smoking on infant mortality came with a strange twist. The LBW babies born to mothers who smoked had lower mortality than the LBW babies of mothers who did *not* smoke. If a baby was born LBW, it seemed an advantage to have a mother who smoked.

These data on the survival of LBW babies provoked a controversy. Yerushalmy was a prominent epidemiologist (and smoker) who defended smoking. One of Yerushalmy's weapons was precisely this observation of better survival among LBW babies born to smokers. He argued that if the survival of these LBW babies was improved by their mothers' smoking, then cigarettes could not be an agent causing them harm. In Yerushalmy's mind, the LBW paradox called into question the causal role of maternal smoking on infant mortality as a whole.[15]

Brian MacMahon rebutted Yerushalmy with a novel argument. MacMahon proposed that a mother's smoking lowered birthweight without affecting the baby's risk.[16] If an exposed baby is smaller but has no corresponding change in its capacity to survive, then the exposed baby's mortality at its new (lighter) weight would be the same as an unexposed baby at the heavier weight. In other words, the smaller infant of a smoking mother might have better survival than other babies at the same weight because the exposed baby still carried the lower risk of its (unachieved) heavier weight. (This argument is discussed in greater detail below.)

MacMahon's insight was subtle, profound, and unappreciated. It was not his argument that ultimately defeated Yerushalmy, but rather the sheer weight of evidence against smoking. Meanwhile, the LBW paradox among the small babies of smoking mothers persists to this day.

The LBW paradox is not unique to smoking. It is also found among babies born at high altitude compared to low altitude,[2] African-American babies compared with white babies,[17] twins compared with singletons,[18] US births compared with Norwegian births[19] and many other examples. Researchers have tried to explain this paradox as due to confounding by gestational age, physiological differences, or specific diseases, but no explanation has withstood testing (e.g. ref. 20). As MacMahon realized, the answer does not lie in confounding but rather in the deeper assumptions brought to the analysis of birthweight. In order to lay the groundwork for re-examining these assumptions, the

following section describes the basic epidemiological features of birthweight—features often neglected in the emphasis on birthweight as a dichotomy

The Epidemiology of Birthweight

Frequency distribution

The frequency distribution of birthweight is strikingly Normal (or bell-shaped), with an extended lower tail. The bar graph in Figure 1 shows the observed distribution of weights for 400 000 births. The Normal component of the birthweight distribution, called the 'predominant' distribution is indicated by the solid line.[14] The predominant distribution (defined by its mean and standard deviation [SD]) comprises the vast majority of births. The remainder of the birthweight distribution is the 'residual' distribution. This residual comprises all births in the lower tail of the curve that falls outside the predominant distribution. In a typical population, 2 to 5% of births are in the residual distribution. Special statistical methods are needed to estimate the predominant and residual distributions (see below). A small excess of large births is less often found in the upper tail of the birthweight distribution. (Methods have been developed to assess both tails of the distribution simultaneously,[21] although a residual in the upper tail has little impact on infant mortality.)

Biological interpretation

The predominant distribution corresponds closely to the birthweight distribution of term births (⩾37 completed weeks of gestation, counting from the last menstrual period). This can be demonstrated in any large data set—the empirical distribution of term births alone is almost purely Normal, with a mean and SD closely approximated by the predominant distribution of all births.[14] (Thirty-seven weeks is admittedly an arbitrary definition of 'term births'.[22] Even so, the Normality of the distribution of term birthweights is robust against modest adjustments in the definition of 'term'.)

It follows that virtually all births in the residual distribution are preterm. However, not all preterm births are in the residual distribution—just the small ones, which also happen to be the ones at highest risk. Populations with a larger per cent of births in the residual distribution would be expected to have a greater number of small preterm births.

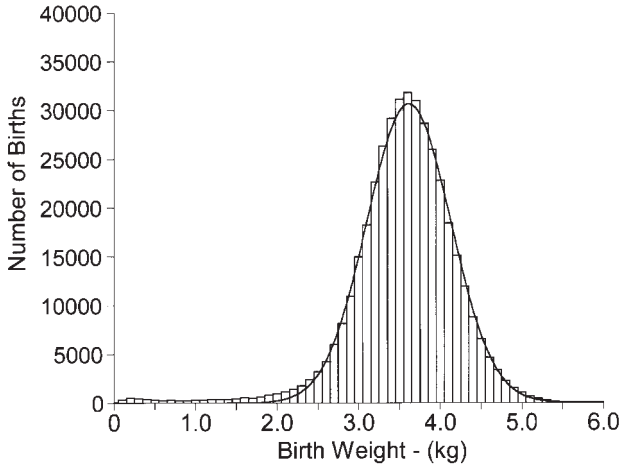

Figure 1 Distribution of birthweights for 405 676 live and still births, Norway, 1992–1998

Thus, the predominant distribution and the residual distribution of birthweight provide indirect information about aspects of gestational age without actually requiring gestational-age data. The predominant distribution closely approximates the weight distribution of term births. The residual distribution estimates the per cent of births that are small and preterm. No other approach to birthweight (certainly not a fixed criterion such as 2500 g) provides this glimpse into a population's gestational-age characteristics. A statistical programme for estimating these two components of the birthweight distribution is available on the internet in a user-friendly format (http://eb.niehs.nih.gov/bwt).

Independence of the two components

The predominant and residual distributions of birthweight are independent of one another. An exposure that affects fetal growth does not necessarily affect the risk of preterm delivery. (The mean of the predominant distribution can change without affecting the per cent of births in the residual distribution.) Conversely, a factor that increases the risk of preterm delivery would not necessarily change the average weight of babies delivered at term. (The per cent in the residual distribution can change without affecting the predominant distribution.) In order to understand birthweight as an epidemiological endpoint, it is essential to grasp this functional independence of the two components of the birthweight distribution.

Implications for infant mortality

When comparing populations of births, a difference in the per cent in the residual suggests a difference in the per cent of small preterm births. Since these are the very babies at highest risk, a population with more babies in the residual distribution will have higher infant mortality (all else being equal).

In contrast, if two populations of babies have different predominant distributions, there is no predictable difference in their infant mortality. Populations with lighter babies do not necessarily have worse mortality. For example, the predominant distribution of Mexican-American babies is shifted to lower weights compared to US white babies, but Mexican-American babies have the better overall survival.[13,23] The mean or SD of the predominant distribution are not reliable indicators of infant mortality.

Reconsidering LBW

How do the two components of the birthweight distribution relate to LBW? Babies less than 2500 g include the whole residual distribution plus the lower tail of the predominant distribution (Figure 1). An increase in residual births (which suggests a health problem) will increase the per cent of LBW. However, the per cent LBW also increases with a decrease in the mean of the predominant distribution, or with an increase in the SD. Such changes in the weight distribution of term births may or may not be associated with changes in mortality. This is why, on a population level, the per cent of LBW is an unreliable marker of perinatal risk.

Birthweight-specific mortality

Birthweight by itself would not have caught the attention of epidemiologists were it not for its association with infant mortality. The relation of mortality to birthweight has a highly

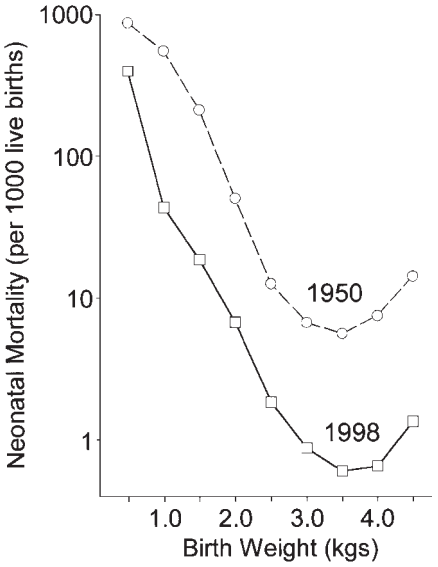

Figure 2 Weight-specific neonatal mortality, USA, 1950 and 1998

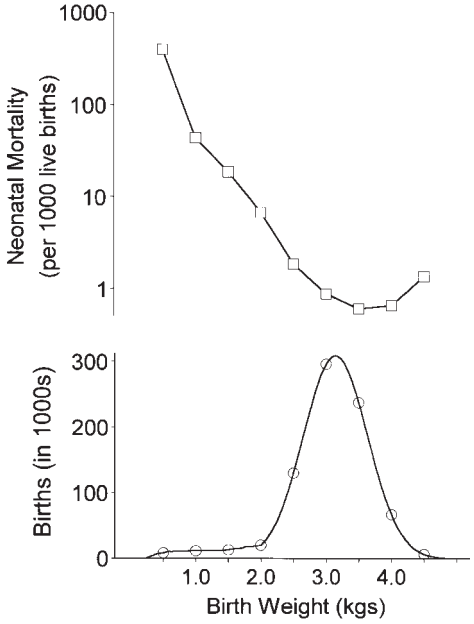

Figure 3 Weight-specific neonatal mortality and the distribution of weights for live births, USA, 1998

distinctive pattern (Figure 2). Mortality ranges more than 100-fold across the spectrum of birthweights. (The Figure shows mortality on a log scale in order to accommodate this huge range.) The reverse-J pattern of weight-specific mortality is found in all populations, and occurs with fetal mortality (stillbirths) and with neonatal or infant mortality.[1] While high mortality among small babies is one of the chief justifications for studying LBW, note the continuous rise of mortality with lower weight. The mortality curve provides no particular justification for 2500 g as the criterion for risk.

Stability of the mortality curve

One fundamental aspect of birthweight-specific mortality is the constancy of its shape. US neonatal mortality fell 75% between 1950 and 1998 (from 20 to 5 per 1000) with no change in the basic shape of the curve (Figure 2).[24,25] This constancy in the shape of this curve over time may be surprising, since much of the improvement in US infant survival over the past 50 years is assumed to be due to better medical care for very small babies.[26] The absolute decline in mortality has indeed been greatest among small infants. However, the *relative* decline in mortality has been fairly uniform across all birthweights (a constant distance on the log scale), with least change at the smallest weights. The general contrast seen between these two mortality curves is typical. The crucial difference in birthweight-specific mortality between any two groups is usually the height of the mortality curves, rather than their shape.[1]

Mean weight and optimum weight

One essential feature of weight-specific mortality is not observable in Figure 2. This feature becomes apparent only when weight-specific mortality is considered in relation to the distribution of birthweights from which the rates are derived (Figure 3). Mean birthweight is several hundred grams lower than optimum birthweight (i.e. the weight with lowest mortality).[27] This difference is maintained across populations and over time, so that as the average birthweight varies, so does optimum weight.[28]

Population geneticists have discussed this phenomenon,[28,29] but epidemiologists have been slower to recognize its implications. The fact that optimum weight maintains a fairly constant distance from mean weight suggests that a shift in the birthweight distribution will produce an equivalent shift in the mortality curve. If all else is held constant, such a shift produces no net effect on infant mortality. The following section pursues this in detail.

An Alternative Hypothesis for the Relation of Birthweight to Infant Mortality

Epidemiologists generally assume that since small babies have high risk, there should logically be an increase in infant mortality with a reduction in mean birthweight. This is not necessarily true.

The effect of altitude

Infant mortality rates are similar in the US as a whole and in the state of Colorado. Most people in Colorado live at high altitudes, and high altitude produces smaller babies. The shift of Colorado birthweights to lower weights is clearly seen in Figure 4 (reprinted from ref. 2). This Figure also shows the curves of weight-specific mortality for Colorado and the US. The two curves intersect. Mortality rates are lower in Colorado for small babies, and higher for large babies. There is no obvious biological explanation for why small babies should do better in Colorado and larger babies should do worse.

Another interpretation of the intersecting mortality curves is that, as birthweights have shifted to lower weights in Colorado, so has optimum weight (and in fact the whole mortality curve). We can test this interpretation by adjusting the two weight distributions to a standard z-scale (with means set to zero and standard deviations to one). Both sets of weight-specific mortality rates are then placed on this z-scale.

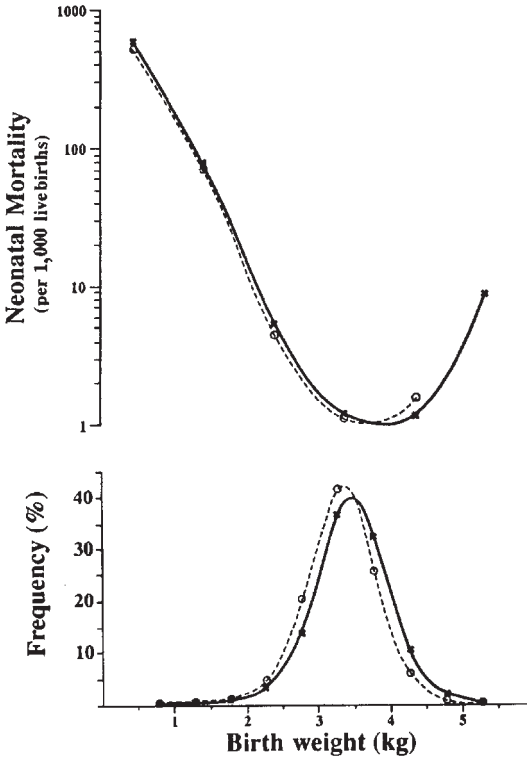

Figure 4 Frequency distribution of birthweight and weight-specific neonatal mortality for Colorado and the United States, 1984. ×—× United States; O—O, Colorado. (Figure reproduced from *Am J Epidemiol* 1993;**137:**1098–104, with permission.)

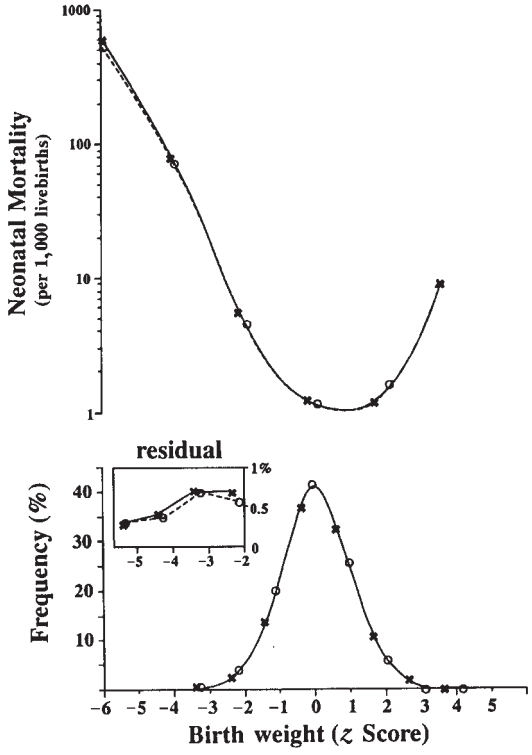

Figure 5 Frequency distribution of birthweight and weight-specific neonatal mortality for Colorado and the United States, 1984, after adjustment to a z scale of birthweight. ×—×, United States; O—O, Colorado. (Figure reproduced from *Am J Epidemiol* 1993;**137:**1098–104, with permission.)

With this adjustment, the two weight distributions correspond nearly exactly, as do the two mortality curves (Figure 5). (The residual distributions are magnified in the inset box for easier inspection.) The simplest explanation for the convergence of mortality curves is that altitude affects birthweight but not mortality. The two mortality curves are essentially the same curve, with the one in Colorado carried along with the shift in birthweight. For babies weighing less than the optimum weight, this shift gives the appearance of lower mortality at any given birthweight. For babies heavier than the optimum weight, the shift gives the appearance of higher mortality. In fact, the birthweight distribution and its accompanying mortality curve have shifted without any change in the survival of individual babies. In this example, fetal growth retardation (on the population level) has no effect on mortality.

We can conclude from this example that the moderate reduction of *in utero* growth does not necessarily increase an individual baby's mortality risk—nor does it increase the number of small babies at higher risk. This might be regarded as a counter-example to Rose's highly-cited thesis that a modest shift in the population mean of a continuous variable (such as blood pressure) will place more individuals into the high-risk group at the extreme.[30] This appears not necessarily to be true for birthweight.

Now imagine a more complicated but plausible scenario. What if a factor decreases birthweight and also increases infant mortality? The same analytical approach can be applied. In the process, we can discover the underlying sense behind the LBW paradox.

The effect of smoking

Mothers who smoke have smaller babies. Their babies, as a whole, have higher infant mortality. If we look at the birthweight and mortality curves for smokers and non-smokers (Figure 6, reprinted from ref. 2), the initial picture is similar to Colorado and the US (Figure 4), if more exaggerated. There are two distinct distributions of birthweight, and the two mortality curves intersect. Small babies have lower mortality if their mothers smoke. This is the paradox by which Yerushalmy defended smoking.

When the picture is adjusted to relative weight (the z-scale), there emerges a new relation between the mortality curves (Figure 7). Mortality with mother's smoking is higher across the whole range of weights. Thus, smoking has two discrete effects. It retards fetal growth, shifting the birthweight distribution (and, as always, the mortality curve). In addition, smoking also shifts the mortality curve upwards, to higher rates. The increased mortality occurs at every adjusted birthweight. In other words, this effect of smoking on weight-specific mortality is independent of birthweight.

The increase of mortality across all weights—*crucial evidence of the harmful effect of smoking on infants*—is initially hidden by the leftward shift of the mortality curve as it follows the birthweight distribution. Small babies of mothers who smoke seem to be at lower risk, when in fact they are at higher risk. This is apparent

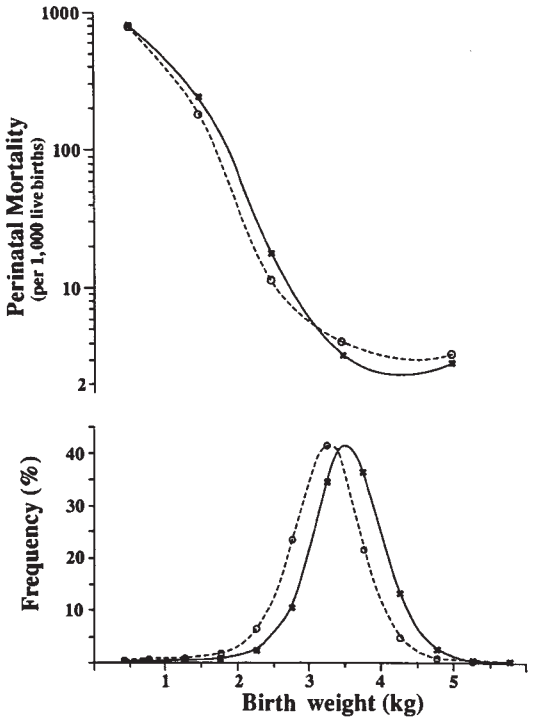

Figure 6 Frequency distribution of birthweight and weight-specific perinatal mortality for infants exposed and unexposed to mothers' smoking: Missouri, 1980–1984. ×—×, non-smokers; O—O, smokers. (Figure reproduced from *Am J Epidemiol* 1993;**137:**1098–104, with permission.)

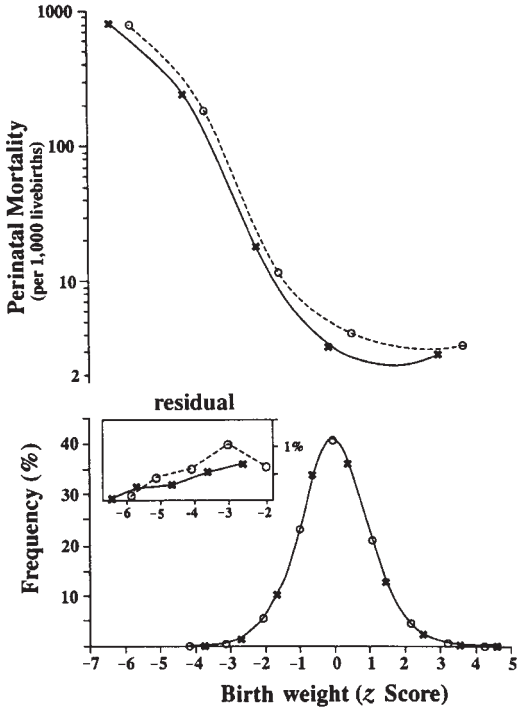

Figure 7 Frequency distribution of birthweight and weight-specific perinatal mortality for infants exposed and unexposed to mothers' smoking, after adjustment to a z scale of birthweight: Missouri, 1980–1984. ×—×, non-smokers; O—O, smokers. (Figure reproduced from *Am J Epidemiol* 1993;**137:**1098–104, with permission.)

on the relative weight scale (the z-scale) but not on the absolute scale. MacMahon anticipated this conclusion when he proposed that the LBW paradox was an artefact due to comparison of absolute weights. Relative weights are needed to uncover the essential relation between smoking and infant mortality. To the extent that smoking increases weight-specific mortality proportionately across all (relative) weights, smoking acts on infant mortality independently of birthweight.

As discussed earlier, the intersection of weight-specific mortality curves is not uncommon. It can be found in nearly any setting where populations have different mean birthweights. In each case, the underlying difference in weight-specific mortality can be revealed by adjustment to a relative scale of birthweight.

Beyond infant mortality

These conclusions extend to any health endpoint associated with birthweight. For example, Liu and colleagues recently published an analysis of cerebral palsy and its association with birthweight in twins and singletons.[31] The LBW paradox was present in their data. Twins had higher rates of cerebral palsy overall, but LBW twins had lower rates of cerebral palsy than LBW singletons. The authors resolved this paradox by adjusting birthweight to the Normal distribution of weight in singletons and twins. After adjustment, the increased risk among twins for cerebral palsy was apparent in every stratum of birthweight.

Implications for Analysis of Birthweight and Related Outcomes

The problem with LBW

The characteristics of the birthweight distribution and its relation to weight-specific mortality provide a foundation for assessing the earlier assumptions about LBW. Is per cent LBW a good surrogate indictor of a population's infant risk? No, because LBW is easily affected by changes in the predominant distribution which are not reliable indicators of risk. Altitude produces more LBW babies, but this does not lead to an increase in infant deaths.[2] Another example is Mexican-American babies. Babies born of Mexican mothers in the US have a predominant distribution of birthweights shifted to lower weights than non-Hispanic whites.[23] This causes Mexican-Americans to have more LBW babies than non-Hispanic whites. However, Mexican-Americans have lower infant mortality.[13] Low birthweight would identify Mexican-Americans as a group at higher risk for infant mortality, but they are not. In this example, the difference in per cent of LBW merely reflects harmless differences in the predominant distribution.

Are LBW births really preventable? Preterm delivery is preventable in principle, and preterm births comprise a major portion of LBW. But what about the lower end of the Normal distribution of births? How can these births be 'prevented'? One option might be to increase the mean or reduce the SD until little of the distribution falls below 2500 g. But if the mortality curve automatically shifts with the birthweight distribution, this strategy is of dubious value. Another alternative would be to change the fundamental Normal distribution of birthweight (for example, by truncating its lower tail). This seems infeasible. Elimination of LBW is neither practical nor necessary in order to achieve the lowest possible rates of infant mortality.

Alternatives to LBW

The arguments above suggest that LBW is muddled as an endpoint, and unreliable as a predictor of population risk. The fact that these uses of LBW are time-honoured is hardly a defence. What alternatives are available? The answer depends on the purpose of the investigator. If the aim is to assess perinatal health through some convenient surrogate, there are several options depending on the type of data available.

When only birthweight is available

If birthweight is the only type of data at hand, the residual distribution should be estimated. The per cent of births in the residual distribution is preferable to LBW as an indicator of perinatal health. The residual provides an estimate of the number of small preterm births—the babies at highest risk. On-line software makes this estimate easy to carry out (http://eb.niehs.nih.gov/bwt).

When birthweight and gestational data are both available

The proportion of preterm births in the population should be examined directly whenever possible. The residual distribution of birthweight is informative, but it is not as good as actual information on preterm delivery. (This of course assumes that the gestational data are of good quality, which is not always the case.) Once the per cent of preterm births is known, the analysis of birthweight can be simplified by restricting the sample to term births. Among term births, the influence of gestational age is minor and can be ignored. The mean and SD of birthweights among term births provide a way to compare fetal growth across groups. The comparison of fetal growth patterns may be interesting in its own right (for example, in understanding the biological effect of a specific exposure), but fetal growth on the population level should not be regarded as a marker of perinatal health.

What about the 'fetal growth curve'?

The pattern of mean birthweights across strata of gestational age has been used to describe the course of intrauterine 'fetal growth'.[32] The assumptions necessary to justify the use of cross-sectional birth data to describe longitudinal growth are dubious at best. At a given gestational age, births are not a random sample of all intrauterine fetuses. This is especially true of births delivered preterm.[33] The use of birth data to describe intrauterine growth patterns is unsound and should be avoided.[34]

What about IUGR (or SGA) as an epidemiologic endpoint?

The use of a weight percentile to define fetal 'growth retardation' has several logical problems. When an external factor (for example, altitude) acts to retard fetal growth, it acts on all babies, not just the small ones. A 9-lb baby can therefore be just as 'growth retarded' as a 5-lb baby when compared with their unaffected weight. Under this scenario, there is no logic in singling out the smallest 10% of babies as the ones who are growth retarded.[35] On a more clinical level, IUGR defined by percentile corresponds poorly with medical signs of fetal growth retardation.[36] Furthermore, IUGR has the unfortunate property of mixing preterm and term births (just as LBW does). If an investigator wishes to summarize intrauterine growth in a population, there is no simpler or more direct endpoint than the mean weight of term births.

The analysis of birthweight becomes even more complicated when birthweight is not the endpoint in itself, but is treated as an intermediate variable. An example is the analysis of infant mortality stratified by birthweight. Such analysis is sometimes done without taking into account the corresponding birthweight distributions.[37,38] This is risky because meaningless differences in weight-specific mortality may be taken as real (as in Figure 4) or important differences may be missed (as in Figure 6). The comparison of US mortality curves in Figure 2 is informative only because the US birthweight distribution has changed so little over the last half-century.

Adjustments of weight-specific mortality can be made using a z-scale, based on the mean and SD of the predominant distribution. A cruder but serviceable method is to compare mortality rates by percentiles of birthweight.[39] The percentile approach may be slightly distorted when study populations differ in their proportion of residual births, but this is probably a minor problem. A method has also been proposed to adjust mortality to a z-scale while controlling for multiple confounding variables.[40]

All these special methods for adjusting to a relative scale of birthweight serve only to underscore one central point. Whatever method is used, excess relative risk tends to be uniform across adjusted birthweights. Despite the huge mortality gradient by birthweight *within* a population, mortality differences *between* populations are generally independent of birthweight.

The unimportance of birthweight

When comparing two populations, the only difference in birthweight that directly affects mortality is a difference in the residual distribution (i.e. a difference in the rate of small preterm births). When infant mortality is higher in one population than another, the mortality difference must be due either to a difference in small preterm births or to differences in weight-specific mortality that are independent of birthweight. This demonstrates the central importance of preterm delivery in infant mortality, and the unimportance of birthweight.

By extension, any analysis of birthweight in relation to associated outcomes must be approached with caution. The most innocent routines of epidemiological analysis are problematic when birthweight is used as an intermediate variable. For example, when analysing infant mortality, epidemiologists often attempt to 'remove' the effects of birthweight by direct or indirect standardization, or by logistic regression. This is presumably done to 'isolate' the mortality effects of factors operating other than through birthweight. As Robins and Greenland have described, this general strategy is unwise.[41] In the specific case of birthweight, the ordinary adjustments of mortality by birthweight implicitly assume that weight-specific differences in mortality are uniform across strata of absolute birthweight. Since weight-specific mortality rates usually intersect under the very conditions that provoke adjustment (i.e. when there are different distributions of birthweight), ordinary birthweight adjustment is nearly always unjustifiable. Furthermore, results of such adjustment have been shown to be biased.[42]

The relation of birthweight to health outcomes in adults

There has been a resurgence of interest in the associations between birthweight and diseases of adulthood—for example, cardiovascular diseases, diabetes, certain cancers, and impairments of hearing or vision.[6,43] It is fascinating to find that, when weight-specific data are available, the risks of later endpoints seem to echo the same reverse-J-shaped pattern seen with infant mortality.[43]

Barker has promulgated the hypothesis that fetal nutrition explains these associations. Fetal nutrition determines fetal growth, fetal growth determines birthweight, and therefore the associations of birthweight with adult diseases demonstrate the impact of fetal nutrition on adult health.[6] However, if (as has been suggested here) the association of birthweight with infant mortality is not causal, there must be similar doubts about birthweight's causal association with diseases in adulthood. Alternative explanations are beginning to emerge, with hypotheses regarding shared genetic mechanisms for fetal growth and later disease.[44–46]

Biological mechanisms that link birthweight to illness or mortality are of great interest, even if they are not causal. Why is infant mortality so strongly related to birthweight, regardless of gestational age? What are the biological underpinnings of the relationship between birthweight and cerebral palsy, or adult hypertension? Perhaps there are metabolism or growth genes that determine fetal size (in some dynamic competition with the maternal system), and that go on to regulate physical development in ways that affect later risk of disease. Such hypotheses offer rich opportunities for further investigation.

In summary, birthweight is strongly associated with a range of health outcomes. These associations have understandably led to an emphasis on birthweight as an epidemiological endpoint in itself. However, this emphasis is misplaced. Birthweight offers little information about population health. Analyses that 'adjust' the effects of birthweight on health outcomes by ordinary means are unsound. Even so, the association of birthweight with so diverse a spectrum of health outcomes is a genuinely fascinating phenomenon. Despite the thousands of papers on birthweight published in past decades, there may be no subject in all of epidemiology more ready for creative—perhaps even revolutionary—insights.

Acknowledgements

The ideas in this commentary have benefited greatly from discussions with my colleagues over the years, most particularly Ian Russell, Pierre Buekens, Rolv Skjaerven, Rolv Terje Lie, and Clarice Weinberg. Matthew Longnecker, Christine Parks, and David Savitz provided helpful comments on the manuscript. Rolv Skjaerven provided the data in Figure 1. D Robert McConnaughey generated the graphs.

References

[1] Wilcox AJ, Russell IT. Birthweight and perinatal mortality: II. On weight-specific mortality. *Int J Epidemiol* 1983;**12:**319–25.

[2] Wilcox AJ. Birthweight and perinatal mortality: the effect of maternal smoking. *Am J Epidemiol* 1993;**137:**1098–104.

[3] Humphrey C, Elford J. Social class differences in infant mortality: the problem of competing hypotheses. *J Biosoc Sci* 1988;**20:**497–504.

[4] Steffensen FH, Sorensen HT, Gillman MW *et al.* Low birthweight and preterm delivery as risk factors for asthma and atopic dermatitis in young adult males. *Epidemiology* 2000;**11:**185–88.

[5] Richards M, Hardy R, Kuh D, Wadsworth MEJ. Birthweight and cognitive function in the British 1946 birth cohort: longitudinal population based study. *Br Med J* 2001;**322:**199–203.

[6] Godfrey KM, Barker DJ. Fetal nutrition and adult disease. *Am J Clin Nutr* 2000;**71(5Suppl.):**1344S–52S.

[7] Mittendorf R, Herschel M, Williams MA, Hibbard JU, Moawad AH, Lee K-S. Reducing the frequency of low birthweight in the United States. *Obstet Gynecol* 1994;**83:**1056–59.

[8] Committee to Study the Prevention of Low Birthweight, Institute of Medicine, National Academy of Sciences, Prevent low birthweight. Washington DC: National Academy Press, 1985.

[9] Public Health Service. Progress toward achieving the 1990 objectives for pregnancy and infant health. *MMWR* 1988;**37:**405–12.

[10] Kliegman RM, Rottman CJ, Behrman RE. Strategies for the prevention of low birthweight. *Am J Obstet Gynecol* 1990;**162:**1073–83.

[11] World Health Organization Expert Committee on Maternal and Child Health. Public health aspects of low birthweight. World Health Organization. Technical Report Series 1961;**217:**3.

[12] Reed DM, Stanley FJ. *The Epidemiology of Prematurity*. Baltimore-Munich: Urban & Schwarzenberg, 1977.

[13] MacDorman MF, Atkinson JO. Infant mortality statistics from the 1997 period linked birth/infant death data set. National vital statistics reports; vol. 47 no. 23. Hyattsville MD: National Center for Health Statistics, 1999.

[14] Wilcox AJ, Russell, IT. Birthweight and perinatal mortality: I. On the frequency distribution of birthweight. *Int J Epidemiol* 1983;**12:**314–18.

[15] Yerushalmy J. The relationship of parents' cigarette smoking to the outcome of pregnancy—implications as to the problem of inferring causation from observed associations. *Am J Epidemiol* 1971;**93:**443–56.

[16] MacMahon B, Alpert M, Salber EJ. Infant weight and parental smoking habits. *Am J Epidemiol* 1966;**82:**247–61

[17] Wilcox AJ, Russell IT. Why small black infants have lower mortality than small white infants: the case for population-specific standards for birthweight. *J Pediatr* 1990;**116:**7–10.

[18] Buekens P, Wilcox AJ. Why do small twins have a lower mortality rate than small singletons? *Am J Obstet Gynecol* 1993;**168:**937–41.

[19] Wilcox AJ, Skjaerven R, Buekens P, Kiely J. Birthweight and perinatal mortality—A comparison of the United States and Norway. *JAMA* 1995;**273:**709–11.

[20] Collins JW Jr, David RJ. Differential survival rates among low-birthweight black and white infants in a tertiary care hospital. *Epidemiology* 1990;**1:**16–20.

[21] Umbach DM, Wilcox AJ. A technique for measuring epidemiologically useful features of birthweight distributions. *Stat Med* 1996;**15:**1333–48.

[22] Papiernik E, Alexander GR, Paneth N. Racial differences in pregnancy duration and its implications for perinatal care. *Med Hypotheses* 1990; **33:**181–86.

[23] Buekens P, Notzon F, Kotelchuck M, Wilcox AJ. Why do Mexican-Americans have few low birthweight infants? *Am J Epidemiol* 2000; **152:**347–51.

[24] US Department of Health, Education and Welfare. Data from the National Vital Statistics System, Series 20, No. 13. A study of infant mortality from linked records. DHEW Publication No. (HSM)72–1056, 1972, Rockville MD.

[25] Matthews TJ, Curtin SC, MacDorman MF. Infant mortality statistics from the 1998 period linked birth/infant death data set. National vital statistics reports; Vol. 48, No. 12, Hyattsville MD, 2000.

[26] Kalter H. Five-decade international trends in the relation of perinatal mortality and congenital malformations: stillbirth and neonatal death compared. *Int J Epidemiol* 1991;**20:**173–79.

[27] Wilcox AJ, Russell IT. Birthweight and perinatal mortality: III. Toward a new method of analysis. *Int J Epidemiol* 1986;**15:**188–96.

[28] Morton NE, Chung CS, Mi MP. *Genetics of Interracial Crosses in Hawaii*. Basel: S Karger Press, 1967, pp.86–92.

[29] Cavalli-Sforza LL, Bodmer WF. *The Genetics of Human Populations*. San Francisco: WH Freeman & Co., 1971, pp.610–14.

[30] Rose G. Sick individuals and sick populations. *Int J Epidemiol* 1985; **14:**32–38.

[31] Liu J-M, Li Z, Lin Q *et al.* Cerebral palsy and multiple births in China. *Int J Epidemiol* 2000;**29:**292–99.

[32] Lubchenko LO, Hansman C, Dressler M, Boyd E. Intrauterine growth as estimated from liveborn birth-weight data at 24 to 42 weeks of gestation. *Pediatrics* 1963;**32:**793–800.

[33] Hediger ML, Scholl TO, Schall JI, Miller LW, Fischer RL. Fetal growth and the etiology of preterm delivery. *Obstet Gynecol* 1995;**85:**175–82.

[34] Wilcox AJ. Birthweight, gestation and the fetal growth curve. *Am J Obstet Gynecol* 1981;**139:**863–67.

[35] Wilcox AJ. Intrauterine growth retardation: beyond birthweight criteria. *Early Hum Dev* 1983;**8:**189–93.

[36] Altman DG, Hytten FE. Intrauterine growth retardation: Let's be clear about it. *Br J Obstet Gynecol* 1989;**96:**1127–32.

[37] National birthweight-specific infant mortality surveillance: preliminary analysis—United States 1980. US Centers for Disease Control. *MMWR* 1986;**35:**269–73.

[38] Erkkola R, Kero P, Seppälä, Grönroos M, Rauramo L. Monitoring perinatal mortality by birthweight specific mortality rates. *Int J Gynaecol Obstet* 1982;**20:**231–35.

[39] Hertz-Picciotto I, Din-Dzietham R. Comparisons of infant mortality using a percentile-based method of standardization for birthweight or gestational age. *Epidemiology* 1998;**1:**61–67.

[40] English PB, Eskenazi B. Reinterpreting the effects of maternal smoking on infant birthweight and perinatal mortality: a multivariate approach to birthweight standardization. *Int J Epidemiol* 1992;**6:** 1097–105.

[41] Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 1992;**3:**143–55.

[42] Wilcox AJ, Russell, IT. Perinatal mortality: standardizing for birthweight is biased. *Am J Epidemiol* 1983;**118:**857–64.

[43] Olsen J, Sørensen HT, Steffensen FH *et al.* The association of indicators of fetal growth with visual acuity and hearing among conscripts. *Epidemiology* 2001;**12:**235–38.

[44] Hattersley AT, Tooke JE. The fetal insulin hypothesis: an alternative explanation of the association of low birthweight with diabetes and vascular disease. *Lancet* 1999;**353:**1789–92.

[45] Stern MP, Bartley M, Duggirala R *et al.* Birthweight and the metabolic syndrome: thrifty phenotype or thrifty genotype? *Diabetes Metab Res* 2000;**16:**88–93.

[46] Lindsay RS, Dabelea D, Roumain J *et al.* Type 2 diabetes and low birthweight—The role of paternal inheritance in the association of low birthweight and diabetes. *Diabetes* 2000;**49:**445–49.

International Journal of Epidemiology 2001;**30**:1241–1243

Commentary: Birthweights and Bell Curves

Richard David

The relationship between an infant's birthweight and survival chances has been recognized for over a 100 years,[1] but in this issue of the *International Journal of Epidemiology* Allen Wilcox has pulled together work from recent decades that brings a clearer understanding of what we know and do not know.[2] His refined analysis of birthweight distributions and how they conform, or fail to conform, to the familiar bell-shaped normal distribution has made possible more informative comparisons between different populations than earlier approaches. Incorrect underlying assumptions or inadequate data weakened much prior work in this field. The point of research in birthweight is to reduce perinatal deaths, and a focus of population comparisons has been to eliminate the gross disparities in outcomes between ethnic groups. The Wilcox model of population birthweight analysis has been faulted by critics as, in effect, perpetuating disparities by lending credence to genetic determinism, an ideology advanced in particularly flagrant terms in Murray and Herrnstein's *Bell Curve*.[3] Is that fair?

Birthweight became a part of the US national standard birth certificate in 1950.[4] Many countries and US states began maintaining computerized birth files that included this important item in the 1950s and we now have archives going back nearly 50 years. Studies of entire populations across generations are becoming common.[5,6] Over this same generation, major changes have taken place in the worlds of politics and science: massive uprisings against racism and colonialism, the transformation of socialism into state (or other) capitalism, the rapid development of scientific disciplines, including, among others, epidemiology and molecular genetics. The elegant mathematics of the normal distribution, first published by De Moivre in 1733, grew out of the study of errors in astronomical measurements, including the work of Galileo.[7] However, in the 21st century as in the 17th and 18th, the models developed by scientists exist in a dialectical give-and-take with wider political and intellectual life in society. Then the central issues might have been whether the earth was the centre of the universe, with implications for the role of the king as centre of the ruling class. Today a debate exists over the genetic meaning or non-meaning of 'race', with implications for the role of relations of dominance on the national and international level. Despite the protests of scientists to the contrary,[8,9] science remains part of the culture that spawns it. This implies social and political responsibilities for scientists.

How does the analysis of birthweight distributions relate to the social conflicts of the past 25 years? A central theme in Allen

Cook County Hospital, University of Illinois at Chicago, Chicago, IL, USA.

Wilcox's work over this period is the contradiction between birthweight as a powerful predictor of mortality in populations and birthweight as a variable lacking any logical causal link to life and death.[10] Mere mass tells the biologist little about an organism's survival potential. Hummingbirds and blue whales have their own adaptations, and size alone, although defining one dimension of diversity, does not imply value. For a given species, however, size measurements can offer hints about past nutrition or present vigour. Size at birth offers additional and special information because it tells us not only the final end-point of growth but may indicate an untimely interruption of that growth *in utero*—by preterm birth. This is the main reason why those who care for sick infants or make policy to promote perinatal health concern themselves with birthweight.

Obviously simply studying gestational age would be more straightforward. After all, most life-threatening newborn conditions result from shortened time *in utero*. For example, respiratory distress syndrome, the leading single pathological entity leading to early newborn death, results from birth prior to the biochemical maturation of the lung.[11,12] However, since the timing of ovulation and conception—or even of the last menses before conception—is often in doubt, studies based on gestational age end up excluding or misclassifying around 5% of births.[13–15] Perhaps a missing data rate of a few per cent seems trivial in a dataset of millions, but the amount of missing and misleading gestational age data is about 50 times more than the amount of missing birthweight data, which is less than 0.1%.[13] Both must be considered in the light of mortality rates in the range 0.5–2%. Clearly, missing data on the wrong births could have a major impact on the outcome measure.

Indeed, missing data on the wrong births is exactly what happens. That is because information on gestational age is not absent or erroneous in a random subset of the population. Those women who have fewer years of education or give birth in their teens or receive inadequate prenatal care are also more likely to have incorrect gestational ages (or none at all) recorded on their infants' birth certificates.[13] These are exactly the babies at greatest risk for death. Having poor prenatal care and limited education are markers for social and economic stressors such as exposure to racism. Racism in turn is associated with preterm birth.[16,17] The oppressed segments of the population are least well tracked for problematic variables like gestational age. Science fails again to escape the bonds of class society. But birthweight is collected accurately for almost every in-hospital birth. Could this offer the escape route for epidemiologists who want to do useful and unbiased analyses without having to first establish equity in health care?

The approach seemed promising, but investigators needed a way to use birthweight, with its the clean and complete data, to draw inferences concerning the rate of preterm birth in populations. The use of the 2500-g cut-point is crude at best. The logical uncoupling of distinct components of the birthweight distribution, in analyses first published by Wilcox and Ian Russell in this Journal in 1983,[18] provided researchers with more powerful conceptual tools for understanding birthweight and mortality. The independence of mean birthweights and preterm birth rates (as reflected in the 'residual' distribution) has been supported by the work of numerous investigators studying different populations in various countries[15,19–23] or the same population over time.[6] Some have duplicated the Wilcox-Russell analysis while others modified it slightly, such as by use of z-scores for birthweight.

One aspect of this useful approach remains problematic. The association of mortality most strongly with the proportion of residual births could imply that the other births are 'normal' in the physiological sense as well as the statistical (i.e. Gaussian) sense. At a population level this has been interpreted to mean that some groups of babies are just 'supposed to be small'. As long as we are discussing individuals living at different altitudes, this is not a problem, but when we start talking about infants of different ethnic groups, concerns develop that real social problems may be defined away rather than vigorously opposed. This is the source of concerns about implied genetic determinism.[3] In the context of a society whose dominant elements justify their positions by arguing the genetic inferiority of those they dominate, it is hard to be neutral. In the pursuit of 'pure science' a well-meaning investigator may be perceived as—and may be—aiding and abetting a social order he abhors.

In fact, evidence indicates that 'normal' birthweight can change for a given ethnic group over time.[6] In the US, white and African American infants both showed increases in mean birthweight over a generation, but not in equal amounts. Black infants did worse, paralleling their disadvantaged position in the American racial hierarchy. Moreover, secular change in births at the extreme low weight end of the curve actually increased for black infants, a trend opposite to their improving mean birthweight, and a change not experienced by whites.[6]

Basically Wilcox is scientifically right: chasing the mean can be misleading, even if most of the time the mode and the tail of the birthweight curve move in tandem. Keeping these outcome parameters distinct will produce improved analyses. The most efficient approach to eliminating racial disparity in perinatal deaths will be targeting the causes for excess preterm births. Many useful studies remain to be done to assess the impact of the social, psychological and environmental factors that underlie preterm birth, and use of a z-score approach will contribute to their clarity and precision. However, in applying an analytical technique to real people, to people living in real societies, to people experiencing births and deaths in their families, the scientist must always proceed with an eye on the social, historical and political context in which those real people—and the scientist—find themselves.

References

[1] Cone TE. Perspectives in neonatology. In: Smith GF, Vidyasager D (eds). *Historical Review and Recent Advances in Neonatal and Perinatal Medicine*. Chicago: Mead Johnson Nutritional Division, 1983.

[2] Wilcox AJ. Reformulations: on the importance—and the unimportance—of birthweight. *Int J Epidemiol* 2001;**30:**1233–41.

[3] Muntaner C, Nieto FJ, O'Campo P. The Bell Curve: on race, social class, and epidemiologic research. *Am J Epidemiol* 1996;**144:**531–36.

[4] US Department of Health, Education and Welfare. *Vital Statistics of the United States, Vol. I.* Washington: US Government Printing Office, 1954, p.20.

[5] Barker D. Fetal origins of coronary heart disease. *Br Med J* 1995;**311:**171–74.

[6] Chike-Obi U, David RJ, Coutinho R, Wu S. Birth weight has increased over a generation. *Am J Epidemiol* 1996;**144:**563–69.

[7] Patel JK, Read CB. *Handbook of the Normal Distribution*. New York: Marcel Dekker, Inc., 1996.

[8] Rothman KJ, Adami HO, Trichopoulos D. Should the mission of epidemiology include the eradication of poverty? *Lancet* 1998;**352:** 810–13.

[9] Zielhuis GA, Kiemeney LALM. Social epidemiology? No way. *Int J Epidemiol* 2001;**30:**43–44.

[10] Wilcox AJ. *Birthweight and Perinatal Mortality* [Dissertation, University of North Carolina]. Ann Arbor, MI: University Microfilms International, 1979.

[11] Binkin NJ, Rust KR, Williams RL. Racial differences in neonatal mortality: what causes of death explain the gap? *Am J Dis Child* 1988; **142:**434–40.

[12] Notter RH. *Lung Surfactants: Basic Science and Clinical Applications*. New York: Marcel Dekker, Inc., 2000.

[13] David RJ. The quality and completeness of birthweight and gestational age data in computerized birth files. *Am J Public Health* 1980;**70:**964–73.

[14] Mustafa G, David RJ. Comparative accuracy of clinical estimates versus menstrual gestational age in computerized birth certificates. *Public Health Rep* 2001 (in press).

[15] Buekens P, Wilcox AJ, Kiely J, Masuy-Stroobant G. Birthweight, preterm births and neonatal mortality in Belgium and the United States. *Paediatr Perinat Epidemiol* 1995;**9:**273–80.

[16] David RJ, Collins JW. Bad outcomes in black babies: Race or racism? *Ethnicity Dis* 1991;**1:**236–44.

[17] Collins JW, David RJ, Symons R, Handler A, Wall SN, Dwyer L. Low-income African-American mothers' perception of exposure to racial discrimination and infant birth weight. *Epidemiology* 2000; **11:**337–39.

[18] Wilcox AJ, Russell IT. Birthweight and perinatal mortality. I. On the frequency distribution of birthweight. *Int J Epidemiol* 1983; **12:**314–18.

[19] Chen R, Wax Y, Lusky A, Toppleberg G, Barell V. A criterion for a standardized definition of low birthweight. *Int J Epidemiol* 1991;**20:** 180–86.

[20] Carlson E, Hoem JM. Low-weight neonatal survival paradox in the Czech Republic. *Am J Epidemiol* 1999;**149:**447–53.

[21] Wilcox A, Skjaerven R, Buekens P, Kiely J. Birth weight and perinatal mortality—a comparison of the United States and Norway. *JAMA* 1995;**273:**709–11.

[22] Olsen SF, Olsen J. A birth-weight adjusted comparison of perinatal mortality in the Faroe Islands and Denmark. *Scand J Soc Med* 1994; **22:**219–24.

[23] Gruenwald P. Fetal growth as an indicator of socioeconomic change. *Public Health Rep* 1968;**83:**867–72.

 International Journal of Epidemiology 2001;**30**:1243–1244

Commentary: When brilliant insights lead astray

Irva Hertz-Picciotto

Wilcox and colleagues have published numerous papers dissecting the relation of birthweight to neonatal mortality from an entirely new angle.[1–6] When I first read these papers in the early 1980s, I found them inspiring. First, they demonstrated that the birthweight distribution consists of a main portion, which has a Gaussian distribution, and an extra residual tail on the left.

Second, they showed how a simple transformation of birthweights to within-population z-scores solved the so-called paradox, whereby certain groups known to have higher neonatal mortality (e.g. Blacks in the US or immigrants in the UK[7]) were found to have, contrary to expectation, better outcomes at low birthweight than Caucasians/Europeans in those countries. In other words, the mortality curves for Whites and Blacks (or immigrants) crossed. Since most deaths among neonates occur in the far more numerous normal birthweight babies, these disadvantaged groups still had higher overall mortality rates. Wilcox discovered that population-specific z-scores for birthweight eliminated the crossover: on the transformed scale, the disadvantaged group had higher neonatal death rates at all weights.

Similar findings were reported using a transformation based on ranks,[8] but the crossover remained on a hybrid scale using absolute deviations from population-specific medians.[9] Translated into epidemiological terminology, each of the purely 'relative' transformations eliminated the effect modification by birthweight,[8,10] and thereby allowed the researcher to treat the problem (under the assumption that birthweight was not an intermediate variable) as one of simple confounding. With a homogeneous (or far less heterogeneous) effect across all birthweights, a summary comparison is more meaningful. Of course, if birthweight is an intermediate on the causal pathway, adjustment as a confounder is not appropriate, and methods such as marginal structural models or structural nested failure time models should be applied.[11,12]

Wilcox provides numerous examples.[6] In the comparison of infants at high altitude versus low, his z-score transformation yields mortality curves that are essentially identical. In other cases, transformation uncovers consistently higher mortality in one group; this scenario holds for US Blacks versus Whites, and for smokers versus non-smokers. The distinction between these

CB #7435, Department of Epidemiology, University of North Carolina, Chapel Hill, NC 27599–7435, USA. E-mail: ihp@unc.edu

two types of relationships—those where the transformed mortality curves are identical versus those where they are not—is crucial. Wilcox acknowledges the distinction, but then loses sight of it.

For example, in 1988, referring to comparisons of US Blacks versus Whites, twins versus singletons, and smokers versus non-smokers, Wilcox *et al*. state that the paradox is 'the artifact of a neutral shift in the mortality curve that accompanies every birthweight shift'.[4] In two of these three examples, however, the shift is not 'neutral'. Wilcox cites MacMahon as having proposed that smoking lowered birthweight but had no effect on mortality risk.[5] If true, one could map each exposed baby back to a counterfactual birthweight, i.e. the weight he/she would have had if exposure had not occurred, and then determine the expected mortality risk for an unexposed baby at that weight. Under MacMahon's hypothesis, this counterfactual risk should correspond to that of the exposed baby. Does it?

Based on Figure 6 in Wilcox's latest paper,[6] the answer is: No! The distance between the birthweight curves is larger than the distance between the mortality curves, i.e. the shift is not neutral. Thus, in Figure 7, after Wilcox's transformation, the mortality curve for infants born to smokers is consistently higher than—not equal to—that of infants born to non-smokers. Wilcox considers MacMahon's insight 'profound,' yet the figures he presents belie MacMahon's hypothesis.

Wilcox takes his argument further, arguing that birthweight is independent of mortality, and hence should not be used as a surrogate adverse outcome. To buttress this argument, he ignores the smoking results (as well as the results of Black versus White comparisons) and relies on cases where 'a shift in the birthweight distribution will produce an equivalent shift in the mortality curve'.[6] To explain away the smoking example, he states 'To the extent that smoking increases weight-specific mortality proportionately across all (relative) weights, smoking acts on infant mortality independently of birthweight.' Again, inspection of his Figure 7 reveals that mortality of smokers' infants is NOT increased proportionately across all weights: the curves do not appear to be equidistant on the log scale. The biggest effect is in the middle of the distribution. The rise in mortality at middle weights is so strong that the reverse J-shape, which Wilcox considers a stable phenomena across populations, is practically replaced by a reverse S-shape, with almost no upturn at higher weights. By Wilcox's own reasoning, the impact of smoking on mortality is not proven to be independent of its effect on birthweight.

One might therefore draw a different conclusion. Perhaps the effects of smoking follow a mixture distribution. Mortality for those at low birthweights is dominated by aspects of suboptimal development that overshadow the impact of smoking (which is nevertheless present, and may operate via birthweight or another pathway). In contrast, babies whose mothers smoked but whose weight still falls within the main portion of the distribution experience markedly higher relative risks for mortality. Whether their reduced birthweight is on the causal pathway or just a marker of higher risk is unclear, but its 'independence' of mortality is not established.

Whereas Wilcox has shown that reduced birthweight is not sufficient in itself to increase mortality, he goes astray when he concludes that 'Birthweight offers little information about population health'. Sometimes it tells us a great deal. Our challenges are to: (1) learn the genetic and environmental influences on birthweight, and (2) distinguish when reduced birthweight is and when it is not prognostic of poorer health and development.

Finally, the easy slide from mortality to morbidity is unwarranted. Even when birthweight is not an indicator of higher mortality risk, does it tell us anything about morbidity? The Wilcox hypothesis of no long-term impact due to lowering of birthweight should be evaluated in studies of birthweight and problems in childhood or beyond: infectious diseases, developmental delays, allergies, etc. Does a within-population transformation give rise to overlapping or distinct distributions of risk for these outcomes? The answer may vary by outcome and by exposure (socioeconomic status, air pollution, nutrition, physical abuse, etc). Until such analyses have produced a clear body of evidence, it is premature to declare that birthweight is unimportant. Already, the evidence suggests that the link between birthweight and health outcomes among infants of smokers is strong, will not disappear by any statistical transformation, is most prominent within the normal weight range, and in the absence of a candidate confounder, likely has a causal component. Regardless of the exposure, if birthweight is on the causal pathway, its use as a surrogate outcome is appropriate and informative.

References

[1] Wilcox AJ, Russell IT. Birthweight and perinatal mortality: I. On the frequency distribution of birthweight. *Int J Epidemiol* 1983; **12:**314–18.

[2] Wilcox AJ, Russell IT. Birthweight and perinatal mortality: II. On weight-specific mortality. *Int J Epidemiol* 1983;**12:**319–25.

[3] Wilcox AJ, Russell IT. Birthweight and perinatal mortality III: Towards a new method of analysis. *Int J Epidemiol* 1986;**15:**188–96.

[4] Skjaerven R, Wilcox AJ, Russell D. Birthweight and perinatal mortality of second births conditional on weight of the first. *Int J Epidemiol* 1988;**17:**830–38.

[5] Wilcox AJ. Birth weight and perinatal mortality: the effect of maternal smoking. *Am J Epidemiol* 1993;**137:**1098–104.

[6] Wilcox AJ. A review: on the importance—and the unimportance—of birthweight. *Int J Epidemiol* 2001;**30:**1233–41.

[7] Marshall T, Mallett R. The influence of racial mix on comparisons of perinatal mortality rate between area health authorities. *Int J Epidemiol* 1980;**9:**255–63.

[8] Hertz-Picciotto I, Din-Dzietham R. Comparisons of infant mortality using a percentile-based method of standardization for birthweight or gestational age. *Epidemiology* 1998;**9:**61–67.

[9] Adams MM. Berg CJ, Rhodes PH, McCarthy BJ. Another look at the Black–White gap in gestation-specific perinatal mortality. *Int J Epidemiol* 1991;**20:**950–57.

[10] Kiely JL, Kleinman JC. Birth-weight-adjusted infant mortality in evaluations of perinatal care: towards a useful summary measure. *Stat Med* 1993;**12:**377–92.

[11] Witteman JC, D'Agostino RB, Stijnen T *et al.* G-estimation of causal effects: isolated systolic hypertension and cardiovascular death in the Framingham Heart Study. *Am J Epidemiol* 1998;**148:**390–401.

[12] Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000;**11:**550–60.

Response: Where do we go from here?

Allen J Wilcox

First, the good news. Neither of my excellent colleagues has defended the study of low birthweight (LBW).[1,2] Hooray. Let us admit LBW is a worn-out approach, and put it away. I am also glad that neither commentator has defended the practice of ordinary adjustment for birthweight in the analysis of infant mortality. The essential assumptions that support such adjustment are nearly always violated. If everyone could agree on these two points, we would spare ourselves a lot of unproductive work.

So what is the bad news? I suppose it is the general confusion over where we go from here. The framework I propose[3] has not been warmly welcomed by the commentators. Dick David implies that my eyes have slipped from the prize—the prize being the reduction of infant mortality. Using race as an example, he suggests that my attempts to steer attention away from birthweight as the source of racial differences in infant mortality are naïve and doomed to misinterpretation. Well, he got that right.

Ian Russell and I published an analysis of Black-White differences in the US[4] which led us to conclude that a high rate of preterm delivery—not lighter mean weight—is the major culprit in the high mortality of African-American babies. Efforts to raise average birthweights (through nutritional programmes, etc.) are unlikely to improve mortality. If we are going for the prize (namely, closing the race gap in mortality), then we should start with the excess of Black preterm births. Racism is likely to be an ingredient in this excess. Data from Collins *et al*. suggest that racial discrimination increases a mother's risk for small preterm births.[5] But much of the US public health community remains committed to birthweight intervention. So whose eyes are on the prize?

Dick David says mean birthweights can be increased—my question is whether they *need* to be. Infant mortality rates have fallen dramatically in recent decades with little or no increase in mean weights. Why couldn't African-Americans be as healthy at birth as European-Americans, regardless of their weight? Dick suggests I am politically incorrect but scientifically right. That's better than the other way around.

Regarding Irva Hertz-Picciotto's comments, I was not clear whether she wants to fine-tune my ideas or toss them out. I will try to address her particular points, saving the big one for last.

Irva says I lose sight of the distinction between identical mortality curves and parallel mortality curves (all this after adjusting for shifts in birthweight). MacMahon[6] did not distinguish between those two—but I do. It is in fact at the foundation of my proposed approach to birthweight.

Another issue Irva raises is whether the adjusted mortality curves are always parallel.

Irva thinks she sees evidence of interaction with birthweight in one of the Figures. She may be right. I have not paid much attention to possible interactions in the smaller data sets because there is so little evidence for it in the huge national data sets. I would be happy to discuss these fine points of adjusted mortality curves—especially if it means we all accept the premises that get us to that point.

In the same spirit, I am open to Irva's suggestion that the relation of birthweight to morbidity may sometimes differ from its relation to mortality. This is a hard question to address because there is not much information on birthweight-specific morbidity. So far, the available data suggest the patterns for weight-specific morbidity and mortality are the same.[3]

Irva's final point is that 'if birthweight is on the causal pathway, its use as a surrogate outcome is appropriate and informative'. This is the crux of the matter. I do *not* think birthweight is on the causal pathway to mortality—at least not for the variables epidemiologists study.

This does not mean that birthweight is an uninteresting endpoint. If racial discrimination were shown to reduce birthweights, it would provide a compelling demonstration of the tangible effects of racism on physical processes. But this would not prove birthweight is the biological mechanism by which racism contributes to infant mortality.

Birthweight is a fascinating variable. Still, when it comes to the study of infant mortality, birthweight is a side issue. Richard David may be right when he says this is politically discomfiting. But if our conventional ways of thinking about birthweight and mortality are unsuccessful, then some new—and uncomfortable—options may be exactly what are needed.

References

[1] David R. Commentary: Birthweights and Bell Curves. *Int J Epidemiol* 2001;**30:**1241–43.

[2] Hertz-Picciotto I. Commentary: When brilliant insights lead astray. *Int J Epidemiol* 2001;**30:**1243–44.

[3] Wilcox AJ. On the importance—and unimportance—of birthweight. *Int J Epidemiol* 2001;**30:**1233–41.

[4] Wilcox AJ, Russell IT. Why small black infants have lower mortality than small white infants: the case for population-specific standards for birthweight. *J Pediatr* 1990;**116:**7–10.

[5] Collins JW Jr, David RJ, Symons R, Handler A, Wall SN, Dwyer L. Low income African-American mothers' perception of exposure to racial discrimination and infant birth weight. *Epidemiology* 2000;**11:**337–39.

[6] MacMahon B, Alpert M, Salber EJ. Infant weight and parental smoking habits. *Am J Epidemiol* 1966;**82:**247–61.