The Analysis of Birth Weight and Infant Mortality: An Alternative Hypothesis

Allen J. Wilcox

Table of Contents

Introduction 3

Why Study Birth weight? 4

A Short History of Low Birth weight 5

The Low Birth weight Paradox 6

The Frequency Distribution of Birth weight 7

Birth weight Specific Mortality 10

The Wilcox-Russell Hypothesis 11

The Analysis of Infant Mortality 18

Beyond Low Birth weight 19

Bibliography 23

The Analysis of Birth Weight and Infant Mortality: An Alternative Hypothesis

Allen J. Wilcox

Birth weight is one of the most commonly studied variables in epidemiology. It is associated with health risks ranging from infant mortality to cardiovascular disease. The usual approach assumes that birth weight is on the causal pathway to whatever health endpoint is of interest. The alternative discussed here is that birth weight is not causally related to health, at least on a population level. If this is true, it has profound implications for the analysis of birth weight, infant mortality or any health endpoint. The material provided on this website has been published as a commentary in IIE (December, 2001) and is reproduced here with permission of Oxford Press.

Website Purpose

- to describe the key features of birth weight as an epidemiologic variable
- to show the problems with low birth weight as a "cause" of infant mortality
- to propose an alternative framework for the relation of birth weight and mortality
- to provide an interactive program for analysis of birth weight distributions

Website Contents

READ	a tutorial discussing the underlying relation of birth weight and infant mortality and its implications for epidemiologic analysis.
ANAYLYZE	an on-line program to analyze birth weight data.
CONTACT	send questions, comments and suggestions to Dr. Wilcox.

Why Study Birth weight?

There are thousands of research papers on birth weight with a hundred more appearing every month. Why is this such a popular topic?

1. **Data are free and abundant**. Birth weight is precisely measurable, recorded by law as part of vital statistics, and available for large populations.
2. **Birth weight is a strong predictor of an individual baby's survival.** In general, the lower the weight, the higher a baby's risk of death.
3. **Groups with lower mean birth weight often have higher infant mortality.** Examples are twins, infants of women who smoke and infants of women with low socioeconomic status.
4. **Low birth weight is associated with poor outcomes later in life.** Asthma, low IQ and hypertension are a few.

Birth weight is usually divided for analysis into "low birth weight" and "normal" birth weight. Behind this simple dichotomy is a complex history and a controversy. You can explore this issue on this website from two directions:

If you're new to the topic of birth weight, you may want to start with The Wilcox-Russell Hypothesis. This sets out a framework for analyzing birth weight. If you have no preconceptions, this is a good introduction. From there, you can proceed to explore the history of other approaches to birth weight and the theoretical basis for the hypothesis.

If you're experienced in analyzing low birth weight, you are going to be challenged to reexamine some of your assumptions. In this case, A Short History of Birth Weight offers a more sensible starting point. From there, you will proceed to The Low Birth Weight Paradox, which sets the stage for an alternative approach. This progression will allow you to see more clearly the connections between the way you're used to thinking about birth weight and the approach proposed here.

To re-iterate- if you're new to the topic, go now to The Wilcox-Russell Hypothesis

If you're experienced in analyzing low birth weight, proceed to A Short History of Low Birth Weight.

I. A Short History of Birth Weight

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

However, not all small babies are premature, and not all premature babies are small. An accumulation of epidemiologic 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. By the 1970s, most researchers were complying, although as late as 1977 a book on LBW was titled *The Epidemiology of Prematurity* . 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 grams 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 percent 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 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 birth weight 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.

1. "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. The sheer strength of this association with mortality is regarded as evidence of its causality.

2. "The percent 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. LBW serves this purpose nicely. Furthermore, under this assumption, the causes of LBW themselves become topics of investigation.

3. "*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.

While these assumptions about LBW are generally accepted, not all aspects of LBW neatly fit into them. For example, groups with a larger percent of LBW babies do not invariably have the greater risk. A well-known example is the comparison of female and male babies. But the most telling contradiction is described in the next section, The Low Birth Weight Paradox.

The Low Birth Weight Paradox

Populations with a higher percent 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.

The Example of smoking

In the 1950's, researchers found that mothers who smoked had smaller babies. By the 1960's, 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. 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. (e.g. Am J Epidemiol 1971;93:445-56)

Brian MacMahon rebutted Yerushalmy with a novel argument. MacMahon proposed that a mother's smoking lowered birth weight without affecting the baby's risk (e.g. Amer J Epidemiol 1966;82:247-61). 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 more completely in The Wilcox-Russell Hypotheses.)

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 (Wilcox 1993), African-American babies compared with white babies (Wilcox 1990), twins compared with singletons (Brekens 1993), US births compared with Norwegian births (Wilcox 1995) and many other examples. Researchers have tried to explain this paradox as due to confounding by gestational age, physiologic differences, or specific diseases, but no explanation has withstood testing. As MacMahon realized, the answer does not lie in confounding but rather in the deeper assumptions brought to the analysis of birth weight. In

order to lay the groundwork for reexamining these assumptions, we must consider the basic epidemiologic features of birth weight – features often neglected in the emphasis on birth weight as a dichotomy.The Frequency Distribution of Birth weight. Start with The Frequency Distribution of Birth Weight.

The Frequency Distribution of Birth Weight

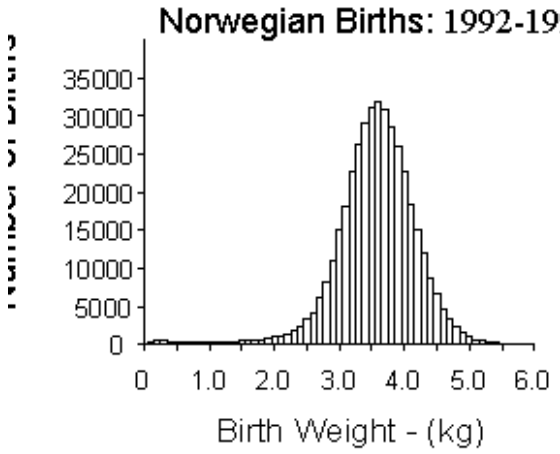


Figure 1: An empirical distribution of 400,000 birth weights (Norway, all births, 1992-1998)

The frequency distribution of birth weight is strikingly Normal (or bell-shaped), with an extended lower tail. The bar graph in Fig. 1 shows the observed distribution of weights for 400,000 births. In Figure 2, the curve superimposed on the bar graph describes the Normal component of the birth weight distribution, called the "predominant" distribution. The predominant distribution (defined by its mean and standard deviation (SD)) comprises the vast majority of births.

The remainder of the birth weight 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. The residual distribution is shown twice in Figure 2, once as the lower tail of the whole distribution, and then enlarged by itself in the bottom panel. 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 birth weight distribution. Methods have been developed to assess both tails of the distribution simultaneously. However, (Umbach 1996), a residual distribution in the upper tail has little impact on infant mortality.

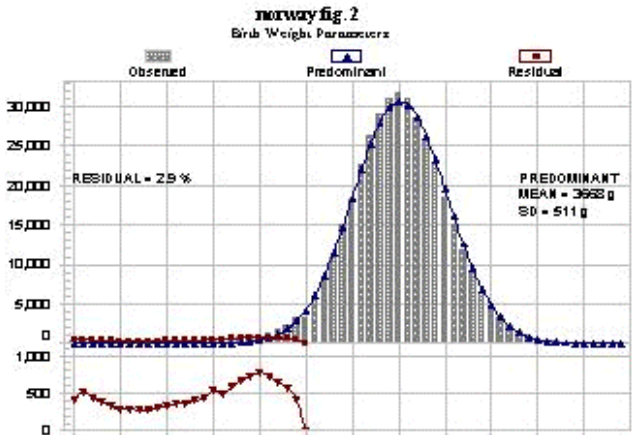


Figure 2: An empirical distribution of 400,000 birth weights, with the estimated predominant and residual distributions

Biological interpretation

Biological interpretation. The predominant distribution corresponds closely to the birth weight distribution of term births (37 or more 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 standard deviation closely approximated by the predominant distribution of all births (Wilcox 1983a). (Thirty-seven weeks is admittedly an arbitrary definition of "term births". The Normality of the distribution of term birth weights remains 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 percent of births in the residual distribution would be expected to have a greater number of small preterm births.

Thus, the predominant distribution and the residual distribution of birth weight 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 percent of births that are small and preterm. No other approach to birth weight (certainly not a fixed criterion such as 2500 grams) provides this glimpse into a population's gestational-age characteristics.

Independence of the two components. The predominant and residual distributions of birth weight 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 percent 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 percent in the residual distribution can change without affecting the predominant distribution). In order to understand birth weight as an epidemiologic endpoint, it is essential to grasp this functional independence of the two components of the birth weight distribution.

Implications for infant mortality. When comparing populations of births, a difference in the percent in the residual suggests a difference in the percent 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. The mean or standard deviation of the predominant distribution are not reliable indicators of infant mortality. (This is discussed more fully in The Wilcox-Russell Hypotheses.)

Reconsidering LBW. How do the two components of the birth weight distribution relate to LBW? Babies less than 2500 grams include the whole residual distribution plus the lower tail of the predominant distribution (Fig. 2). An increase in residual births (which suggests a health problem) will increase the percent of LBW. However, the percent 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 percent of LBW is an unreliable marker of perinatal risk.

Summary

The birth weight distribution tells something about small preterm births, but not by using a simple cut-off of 2500 grams. A more complicated estimation procedure is needed to describe the residual distribution. This website includes a Birth weight Analysis Program that estimates the predominant and residual distributions for any birth weight distribution. But before using this program, please read more about the birth weight story in Birth weight-Specific Mortality.

Birth weight-Specific Mortality

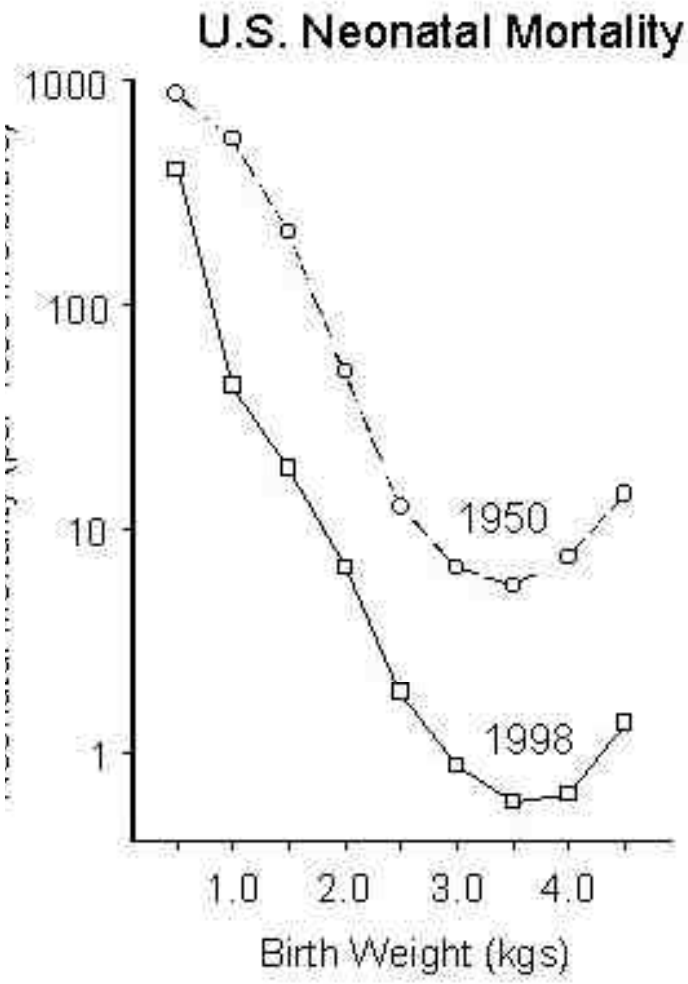


Figure 3: Weight-specific neonatal mortality for the US, 1950 and 1998

Birth weight by itself would not have caught the attention of epidemiologists were it not for its association with infant mortality. The relation of mortality to birth weight has a highly distinctive pattern (Fig. 3). Mortality ranges more than 100-fold across the spectrum of birth weights. (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 (Wilcox 1983b). 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 grams as the criterion for risk.

Stability of the curve.

One fundamental aspect of birth weight-specific mortality is the constancy of its shape. US neonatal mortality fell 75% between 1950 and 1998 (from 20 to 5 per thousand) with no change in the basic shape of the

curve (Fig. 3). 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 fifty years is assumed to be due to better medical care for very small babies. The absolute decline in mortality has indeed been greatest among small infants. However, the relative decline in mortality has been fairly uniform across all birth weights (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 birth-weight-specific mortality between any two groups is usually the height of the mortality curves, rather than their shape (more details in Wilcox 1983b).

There is one more feature of the mortality curve that becomes apparent only when the curve is considered in relation to the birth weight distribution. This feature is the foundation for the next section, The Wilcox-Russell Hypothesis. (If you've already read "The Wilcox-Russell Hypothesis", then go directly to The Analysis of Infant Mortality.)

The Wilcox-Russell Hypothesis

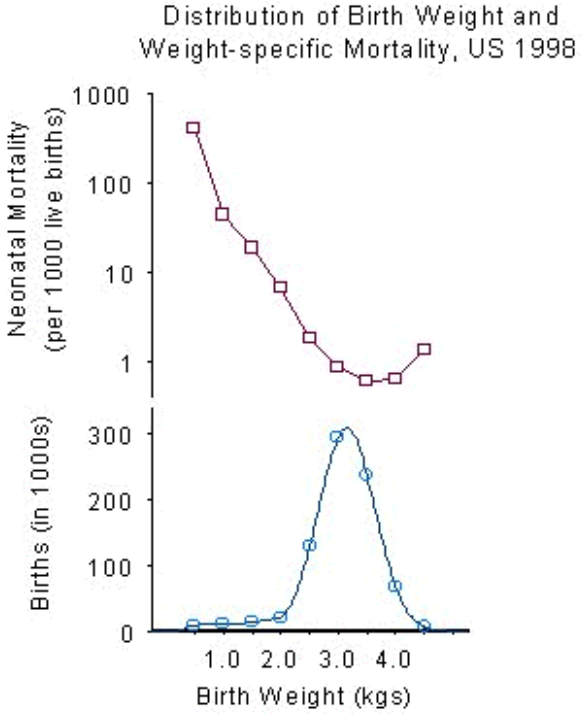
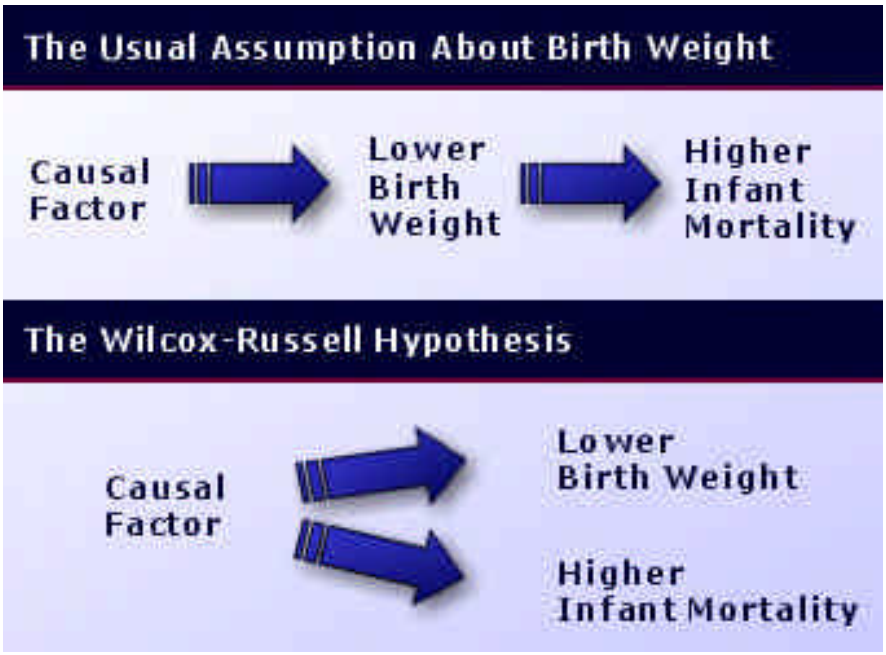


Figure 4: Empirical distribution of birth weight and weight-specific neonatal mortality for the US, 1998

A bell-shaped distribution of birth weight is found in all populations (The Frequency Distribution of Birth weight). Similarly, the curve of weight-specific mortality has the same general shape in all populations (Birth weight-Specific Mortality). When the weight and the mortality curves are taken together, an additional feature emerges. The mean weight is always several hundred grams lower than the optimum weight (the weight with lowest mortality). Just as the average birth weight varies among populations, so does the optimum weight. This relation between optimum and mean weight is the basis for the Wilcox-Russell hypothesis.

The Wilcox-Russell hypothesis. The usual assumption about birth weight is that a change in birth weight directly affects perinatal survival. The Wilcox-Russell hypothesis is that, on a population level, birth weight is not on the causal pathway to mortality. A change in birth weight is often *associated* with a change in perinatal health but it is not *through* the change in weight that the health effect occurs. Weight and mortality can change together because a single factor is affecting them both. But birth weight can also change without an effect on mortality.


The Usual Assumption About Birth Weight
Causal Factor
Lower Birth Weight
Higher Infant Mortality
The Wilcox-Russell Hypothesis
Causal Factor
Lower Birth Weight
Higher Infant Mortality

Support for this hypothesis is found in the fact that, as mean birth weight changes, the weight-specific mortality curve changes by exactly equivalent amounts. Data from the US and Colorado offer a good example.

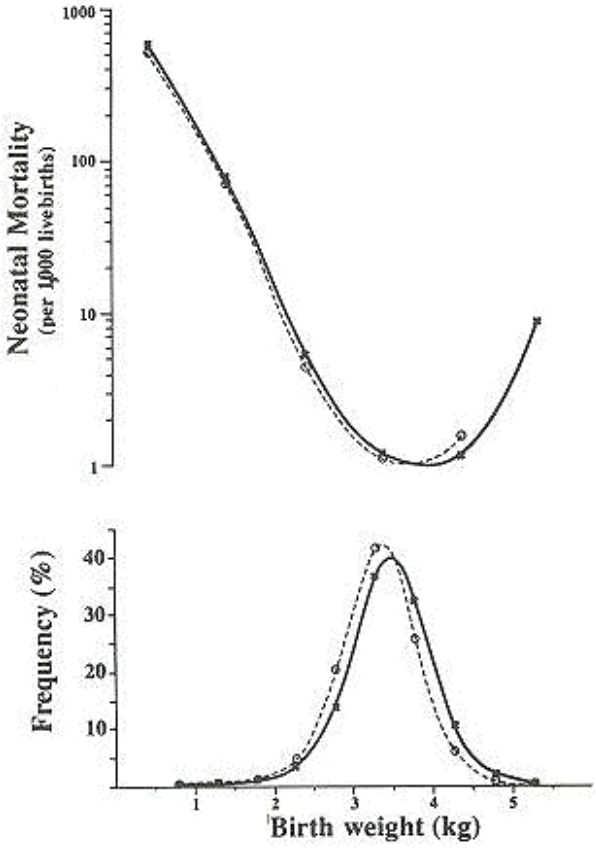


Figure 5: Frequency distributions of birth weight and weight-specific neonatal mortality rates for Colorado and the United States, 1984. x___x, United States; O---O, Colorado. (Figure reproduced from *Am J Epidemiol*, 1993; 137; 1098-1104, with permission.)

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 birth weights to lower weights is clearly seen in Fig. 5.

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 birth weights 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.

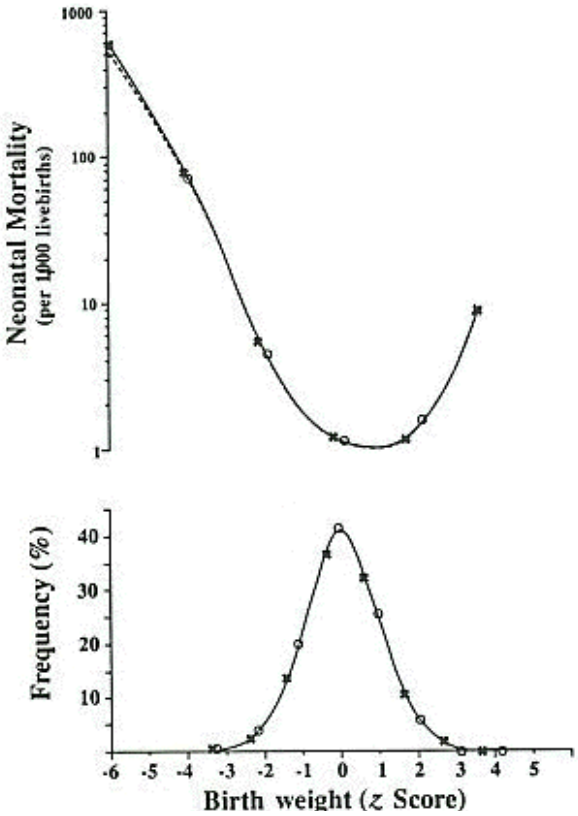


Figure 6: Frequency distributions of birth weight and weight-specific neonatal mortality rates for Colorado and the United States, 1984, after adjustment to a z scale of birth weight. x____x, United States; O---O, Colorado. (Figure reproduced from *Am J Epidemiol*, 1993; 137; 1098-1104, with permission.)

With this adjustment, the two weight distributions correspond nearly exactly, as do the two mortality curves (Fig. 6). The simplest explanation for the convergence of mortality curves is that altitude affects birth weight but not mortality.

The two mortality curves are essentially the same curve, with the one in Colorado carried along with the shift in birth weight. For babies weighing less than the optimum weight, this shift gives the appearance of lower mortality at any given birth weight. For babies heavier than the optimum weight, the shift gives the appearance of higher mortality. In fact, the birth weight distribution and its accompanying mortality curve has 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 Geoffrey 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. This appears not necessarily to be true for the birth weights of term babies.

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

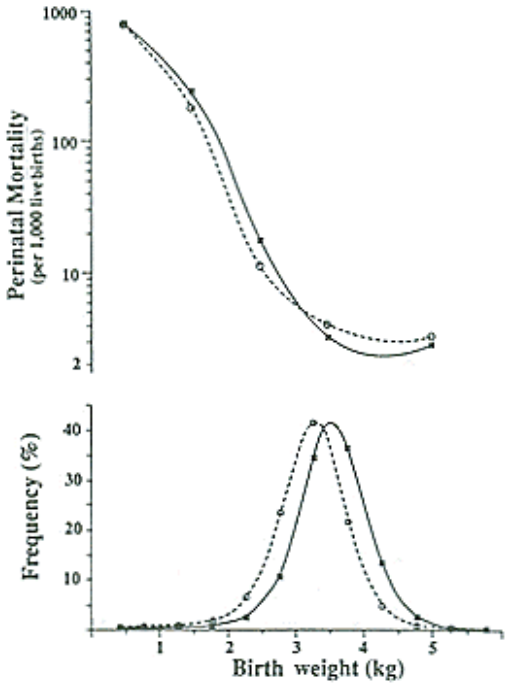


Figure 7: Frequency distributions of birth weight and weight-specific perinatal mortality rates for infants exposed and unexposed to mothers' smoking: Missouri, 1980-1984. x___x, nonsmokers; O---O, smokers. (Figure reproduced from *Am J Epidemiol*, 1993; 137; 1098-1104, with permission.)

The effect of smoking. Mothers who smoke have smaller babies. Their babies have higher infant mortality as a group. If we look at the birth weight and mortality curves for smokers and non-smokers, the initial picture is rather similar to Colorado-US. There are different birth weight distributions and the two mortality curves intersect. Small babies do better if their mothers smoke. This is the paradox with which Yerushalmy defended smoking.

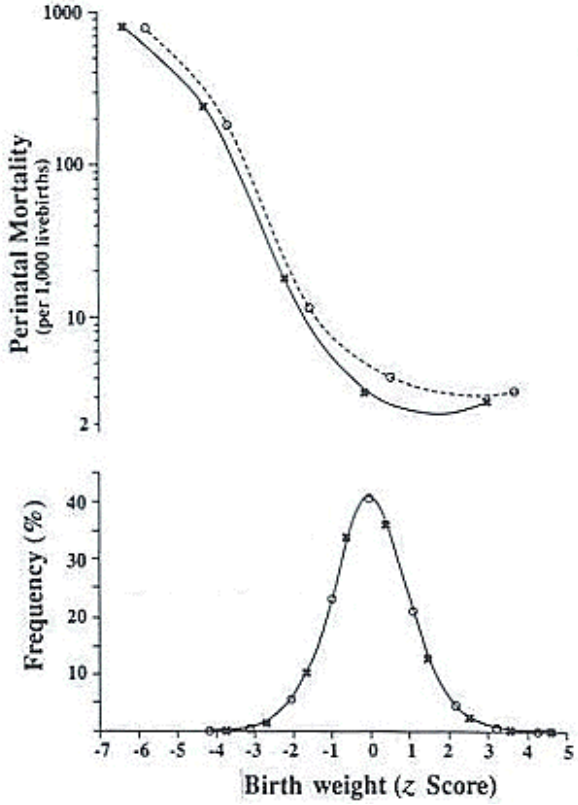


Figure 8: Frequency distributions of birth weight and weight-specific perinatal mortality rates for infants exposed and unexposed to mothers' smoking, after adjustment to a z scale of birth weight: Missouri, 1980-1984. x---x, nonsmokers; O---O, smokers. (Figure reproduced from *Am J Epidemiol*, 1993; 137; 1098-1104, with permission.)

When the picture is adjusted to relative weight (the z-scale), there emerges a new relation between the mortality curves (Fig. 8). 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 birth weight distribution (and, as always, the mortality curve). In addition, smoking also shifts the mortality curve upwards, to higher rates.

In the previous example of altitude, the shift of the birth weight distribution to lower weights was not sufficient to increase infant mortality. In the example of smoking, there is increased mortality that occurs equally at every adjusted birth weight (on a multiplicative scale). In other words, this effect of smoking on weight-specific mortality is independent of birth weight.

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 birth weight distribution. Small babies of mothers who smoke seem to be at lower risk, when in fact they are at higher risk. This is apparent 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 artifact due to comparison of absolute weights. (See the Low Birth Weight Paradox) 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 independent of birth weight.

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 birth weights. In each case, the true difference in weight-specific mortality is revealed after adjustment to a relative scale of birth weight.

If you have not yet read A Short History of Low Birth weight, now would be a good time to do so. If you've already read it, go now to The Analysis of Infant Mortality.

The Analysis of Infant Mortality

If birth weight is not on the causal pathway to infant mortality, then birth weight is neither an explanatory factor nor a potential confounding variable in the analysis of infant mortality. As a general statement, birth weight is not important in the analysis of infant mortality. In contrast, preterm delivery is on the causal pathway to infant mortality. Any analysis of infant mortality should take into account the contribution of preterm delivery. However, gestational age data are often incomplete or of poor quality, which can make it difficult to identify preterm births. The importance of birth weight data, therefore, is to provide inferences about preterm births.

The residual distribution identifies the percent of small, high-risk preterm births in any study population. (A birth weight analysis program is available on this website to estimate the residual distribution.) If you also have data on birth-weight-specific mortality, then the number of deaths among residual births can be calculated by applying the weight-specific mortality rates to the residual distribution. By inference, this provides the number of deaths among small preterm births. These deaths typically amount to half or more of all infant deaths.

After taking account of infant deaths among the small preterm infants, the remaining deaths are all among the predominant distribution. Virtually all of these are term births. Among term births, it is not the distribution of birth weights (which is always Normal, or gaussian) but the height of the mortality curve that determines the number of deaths.

The birth weight distribution of term babies is useful for one thing: it provides information about fetal growth. Keep in mind that the consequences of fetal growth for infant mortality are unpredictable. Even so, fetal growth can be a useful endpoint in itself for assessing the biological effects of nutrition, environmental exposures, etc. The specifics of analyzing fetal growth and birth weight are discussed in the next section (Beyond Low Birth weight).

You're almost done. Beyond Low Birth weight is the final section.

Beyond Low Birth Weight

Problems with LBW

The hypothesis has been proposed that birth weight is not on the causal pathway to infant mortality (The Wilcox-Russell hypothesis). If this hypothesis is correct, we must reexamine some of the basic assumptions about LBW (see A short history of low birth weight).

1. Is percent 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 (Wilcox 1993). Another example is Mexican-American babies. Babies born of Mexican mothers in the US have a predominant distribution of birth weights shifted to lower weights than non-Hispanic whites (Brekens 2000). This causes Mexican-Americans to have more LBW babies than non-Hispanic whites. However, Mexican-Americans have lower infant mortality. LBW would identify Mexican-Americans as a group at higher risk for infant mortality, but they are not.

Difference in percent of LBW may reflect harmless differences in the predominant distribution.

2. 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 2500g. But if the mortality curve automatically shifts with the birth weight distribution, this strategy is of dubious value.

Another alternative would be to change the fundamental Normal distribution of birth weight (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 in the analysis of birth weight

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-honored is hardly a defense.

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.

1. *When only birth weight is available*. If birth weight is the only type of data at hand, the residual distribution should be estimated. The percent 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.

2. *When birth weight and gestational data are both available.* The proportion of preterm births in the population should be examined directly whenever possible. The residual distribution of birth weight 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 percent of preterm births is known, the analysis of birth weight 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 birth weights 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 not a dependable marker of perinatal health.

3. *What about the "fetal growth curve"?* The pattern of mean birth weights across strata of gestational age has been used to describe the course of intrauterine "fetal growth". 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. The use of birth data to describe intrauterine growth patterns is unsound and should be avoided.

4. *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 nine-pound baby can therefore be just as "growth retarded" as a five-pound 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.

On a more clinical level, IUGR defined by percentile corresponds poorly with medical signs of fetal growth retardation. 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 perils of ordinary adjustments by birth weight

The analysis of birth weight becomes even more complicated when birth weight is not the endpoint in itself, but is treated as an intermediate variable. An example is the analysis of infant mortality stratified by birth weight. Such analysis is sometimes done without taking into account the corresponding birth weight distributions. This is risky because meaningless differences in weight-specific mortality may be taken as real (as in Fig 5) or important differences may be missed (as in Fig. 7). The comparison of US mortality curves in Figure 3 is informative only because the US birth weight 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 birth weight. 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 (English 1992).

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

The unimportance of birth weight. When comparing two populations, the only difference in birth weight 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 birth weight. This demonstrates the central importance of preterm delivery in infant mortality, and the unimportance of birth weight.

By extension, any analysis of birth weight in relation to associated outcomes must be approached with caution. The most innocent routines of epidemiologic analysis are problematic when birth weight is used as an intermediate variable. For example, when analyzing infant mortality, epidemiologists often attempt to "remove" the effects of birth weight by direct or indirect standardization, or by logistic regression. This is presumably done to "isolate" the mortality effects of factors operating other than through birth weight.

As Robins and Greenland have described, this general strategy is unwise (Epidemiology 1992; 3:143-155). In the specific case of birth weight, the ordinary adjustments of mortality by birth weight implicitly assume that weight-specific differences in mortality are uniform across strata of absolute birth weight. Since weight-specific mortality rates usually intersect under the very conditions that provoke adjustment (i.e. when there are different distributions of birth weight), ordinary birth weight adjustment is nearly always unjustifiable. Furthermore, results of such adjustment have been shown to be biased (Wilcox 1983c).

The relation of birth weight to later health outcomes

There has been a resurgence of interest in the associations between birth weight and diseases of adulthood – for example, cardiovascular diseases, diabetes, certain cancers, and impairments of hearing or vision. It is

fascinating to find that, when weight-specific data are available, the risks of later endpoints often echo the same reverse-J-shaped pattern seen with infant mortality.

Barker has promulgated the hypothesis that fetal nutrition explains these associations. Fetal nutrition determines fetal growth, fetal growth determines birth weight, and therefore the associations of birth weight with adult diseases demonstrate the impact of fetal nutrition on adult health. However, if (as has been suggested here) the association of birth weight with infant mortality is not causal, there must be similar doubts about birth weight'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.

Biological mechanisms that link birth weight to illness or mortality are of great interest, even if they are not causal. Why is infant mortality so strongly related to birth weight, regardless of gestational age? What are the biological underpinnings of the relationship between birth weight 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.

Coda

In summary, birth weight is strongly associated with a range of health outcomes. These associations have understandably led to an emphasis on birth weight as an epidemiologic endpoint in itself. However, this emphasis is misplaced. Birth weight offers little information about population health.Analyses that "adjust" the effects of birth weight on health outcomes by ordinary means are unsound.

Even so, the association of birth weight with so diverse a spectrum of health outcomes is a genuinely fascinating phenomenon. Despite the thousands of papers on birth weight published in past decades, there may be no subject in all of epidemiology more ready for creative – perhaps even revolutionary – insights.

Comments, suggestions, or questions can be directed to Alan Wilcox. Suggestions for improvements of this website will be incorporated in future versions. Relevant citations for the Bibliography are welcomed.

BIBLIOGRAPHY:
On Birth weight and the Wilcox-Russell Hypothesis

CORE METHODS

Wilcox AJ, Russell, IT. Birth weight and perinatal mortality: I. On the frequency distribution of birth weight. *Int J Epidemiol* 12:314-18, 1983a.

Wilcox AJ, Russell, IT. Birth weight and perinatal mortality: II. On weight-specific mortality. *Int J Epidemiol* 12:319-25, 1983b.

Wilcox AJ, Russell IT. Birth weight and perinatal mortality: III. Toward a new method of analysis. *Int J Epidemiol* 15:188-96, 1986.

OTHER THEORETICAL ASPECTS

Wilcox AJ. Birth weight, gestation and the fetal growth curve. *Am J Obstet Gynecol* 39:863-67, 1981.

Wilcox AJ. Intrauterine growth retardation: beyond birth weight criteria. *Early Hum Dev* 8:189-93, 1983c.

Wilcox AJ, Russell, IT. Perinatal mortality: standardizing for birth weight is biased. *Am J Epidemiol* 118:857-64, 1983d.

Weinberg CR, Wilcox AJ. Reproductive epidemiology. In: Rothman KJ, Greenland S (eds.): *Modern Epidemiology* (2nd Edition), Philadelphia, Little, Brown & Co., 1998, 585-608.

EXTENSIONS

English PB, Eskenazi B. Reinterpreting the effects of maternal smoking on infant birth weight and perinatal mortality: a multivariate approach to birth weight standardization. *Int J Epidemiol*. 6:1097-105, 1992.

Umbach DM, Wilcox AJ. A technique for measuring epidemiologically useful features of birth weight distributions. *Statist Med* 15: 1333-48; 1996.

Hertz-Picciotto I, Din-Dzietham R. Comparisons of infant mortality using a percentile-based method of standardization for birth weight or gestational age. *Epidemiol* 1:61-7, 1998.

Liu J-M, Li Z, Lin Q, Zhao P, Zhao F-L, Hong S-X, Li S. Cerebral palsy and multiple births in China. *Internat J Epidemiol* 29:292-99, 2000.

APPLICATIONS

Skjærven R, Wilcox AJ, Russell D. Birth weight and perinatal mortality of second births conditional on weight of the first. *Int J Epidemiol* 17:830-38, 1988.

Bjerkedal T, Czeizel A, Hosmer DW Jr. Birth weight of single live births and weight specific early neonatal mortality in Hungary and Norway. *Paediatr Perinat Epidemiol*. 1:29-40, 1989.

Wilcox AJ, Russell, IT. Why small black infants have lower mortality than small white infants: the case for population-specific standards for birth weight. *J Pediatr* 116:7-10, 1990.

Wilcox AJ, Skjærven R. Birth weight and perinatal mortality: the effect of gestational age. *Am J Public Health* 82:378-82, 1992.

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

Wilcox AJ. Birth weight and perinatal mortality: the effect of maternal smoking. *Am J Epidemiol* 137:1098-1104, 1993.

Wilcox AJ, Skjærven R, Buekens P, Kiely J. Birth weight and perinatal mortality: A comparison of the United States and Norway. *JAMA* 273:709-11; 1995.

Øyen N, Skjærven R, Little RE, Wilcox AJ. Fetal growth retardation in Sudden Infant Death Symdrome (SIDS) babies and their siblings. *Am J Epidemiol* 142:84-90; 1995.

Buekens P, Wilcox AJ, Kiely J, Masuy-Stroobant G. Birth weight, preterm births and neonatal mortality in Belgium and the United States. *Pediatr Perinat Epidemiol* 9:273-80; 1995.

Skjærven R, Wilcox AJ, Øyen N, Magnus P. Mothers' birth weight and the survival of their offspring. *Brit Med J* 314: 1376-80; 1997.

Buekens P, Notzon F, Kotelchuck M, Wilcox AJ. Why do Mexican-Americans have few low birth weight infants? *Amer J Epidemiol* 152:347-51; 2000.