Wilcox and colleagues have published numerous papers dissecting the relation of birthweight to neonatal mortality from an entirely new angle.\(^8\) 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,\(^9\) 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.\(^5\) 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

---

**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,\(^5\) 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.\(^5\) 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
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’. 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. 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, 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’. 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