The renowned statistician George P. Box famously said that all models are wrong, but some are useful. Far from an indictment of statistical models, Box’s statement can be taken to mean that even when complex realities are not exactly represented by simple fitted models, much can be learned. The paper by Basso et al. (1) in this issue of the Journal provides an opportunity to consider the costs and benefits that arise from the simplification necessary for generating statistical models of complex biologic processes. Considering the relation among birth weight, mortality, and third factors, Basso et al. postulate that birth weight is not itself on the causal path to mortality; rather, the relation between birth weight and mortality might be explained by a confounding factor. The authors conclude that, to produce the observed inverse J shape of the birth-weight-specific mortality curve, the putative confounding factors (matrix $X = (X_1$ and $X_2)$) must be very rare and have very large effects.

In reducing complex situations to simple models, assumptions are made, especially when modeling a biologic process. For example, parametric models make assumptions regarding distributions. The flexibility of the models is limited by that of the assumptions on which it depends. Basso et al.’s model makes the following assumptions: 1) birth weight follows a Gaussian distribution, 2) there is a uniform effect of the confounding factors $X_1$ and $X_2$ on mortality, 3) birth weight does not cause neonatal mortality, and 4) there is no interaction between factors $X_1$ and $X_2$ and birth weight. Notably, some of these assumptions are interrelated, and a change in one might affect the others.

As previously stated, consideration of the potential limitations of this model requires further attention to the assumptions on which it is based. Let us review each of these assumptions regarding their substance, the possible impact on the findings if they are violated, and whether they seem reasonable.

ASSUMPTIONS

Gaussian distribution for birth weight

This assumption requires that birth weight follow a Normal distribution at all strata of the confounding factor. The hypothetical birth weights in the left tail of this distribution may be regarded with some skepticism because of their questionable compatibility with viability. If the left tail is indeed truncated, the Gaussian birth-weight assumption will be violated. Moreover, a confounding factor might increase the proportion of low-birth-weight babies without shifting the whole birth-weight distribution, resulting in a skewed distribution within that stratum. However, given the relatively low prevalence of fetal-growth-restricted babies, major deviations from a Gaussian distribution are unusual in real life. Additionally, a small violation of this assumption will have little impact on the shape of the overall association between birth weight and neonatal mortality.

Uniform effect of risk factors

This assumption states that all babies exposed to the putative confounding factor $X$ are assumed to have identical shifts in birth weight and identical elevation of their mortality risk. In their paper, Basso et al. acknowledge that this assumption is unlikely to be true. However, a modification of this assumption seems to entail minor changes on the noncausal link between birth weight and neonatal mortality. The authors refined the model by substituting distributions for the constant effects and obtained similar results.
Noncausal effect of birth weight on neonatal mortality

This assumption states that neonatal mortality is independent of birth weight conditional on the confounding factor that links them or, stated more simply, that birth weight in no way has any effect on neonatal mortality. In probability terms, this assumption implies conditional independence and can be rewritten as

\[
\Pr(NM = 0 \mid BW = bw, X = x) = \Pr(NM = 0 \mid X = x) = f(x), \quad (1)
\]

where \( NM = 0 \) indicates neonatal mortality status, \( BW \) is birth weight and \( X \) is the unmeasured confounder, and \( f(x) \) is a function that depends only on a value \( x \) of factor \( X \). Departure from this assumption could lead to different conclusions. As soon as the model allows for a causal effect of birth weight on neonatal mortality, the risk factors (\( X_1 \) and \( X_2 \), following the notation of Basso et al.) no longer need to be rare or have large effects.

Considering equation 1, what are the conditions and characteristics of the putative confounding factor, \( X \), that would give rise to the observed shape of the curve between birth weight and neonatal mortality? Such a factor must comply with the following equality that depicts the joint probability of neonatal mortality and birth weight so that it results in the inverse J-shaped curves, with form

\[
\Pr(NM = 0, BW = bw) = \int f(x)\Pr(BW = bw \mid X = x)F_X(dx). \quad (2)
\]

It is worth noting that, from such conditions, the strength of the effect of the factor \( X \) on birth weight plays a crucial role in determining the shape of the crude association between birth weight and neonatal mortality under conditional independence assumptions. Given these circumstances, Basso et al. show that the confounding factor \( X \) cannot be smoking, or any other factor with similar properties, because it neither is sufficiently rare nor conveys sufficient risk to satisfy the conditions set forth by equation 2.

However, if one allows for a departure from the conditional independence assumption, the added flexibility allows for a very large range of circumstances and attendant conclusions. When a causal effect between birth weight and neonatal mortality is assumed to be possible, the unmeasured confounder \( X \) is not constrained to be rare or have large effects, and the following applies:

\[
f(bw, x) = \Pr(NM = 0 \mid BW = bw, X = x) \neq \Pr(NM = 0 \mid X = x). \quad (3)
\]

As a result, the equality in equation 2 becomes

\[
\Pr(NM = 0, BW = bw) = \int f(x, bw)\Pr(BW = bw \mid X = x)F_X(dx). \quad (4)
\]

Adding birth weight to function \( f \) creates a large number of alternative routes to solving the equality represented by equation 4. Specifically, nonlinear relations and/or interactions could lead to the observed shape of the birth weight–neonatal mortality curve. Actually, under the relaxed assumption, an infinity of models is possible, including that in which the J-shaped pattern is completely explained by the effect of birth weight on mortality.
Lack of interaction between factors $X_1$ or $X_2$ and birth weight

In reaching their conclusions, Basso et al. assume no interactive effect between risk factors and fetal growth on neonatal mortality. However, in seeking explanations for the observed shape of the birth-weight curves, there are alternatives that merit consideration as well. We simulated an interactive effect between a common factor and fetal growth. To do so, we had to relax another model assumption by allowing a causal effect of fetal growth on neonatal mortality. We were able to generate curves that were similar to the empirically observed data, as shown in figure 1. This provides an alternative to the proposed rare exposure with extreme effects on mortality.

There is biologic plausibility to both scenarios—that of no direct effect from birth weight to fetal growth and that of at least a small direct effect with or without interactions. In addition, both scenarios have similar practical implications. Under the one presented by Basso et al., the putative “rare” exposures will be difficult to uncover, although perhaps we could identify them among the causes of death attributed to low-birth-weight babies on death certificates. Unfortunately, the same is true under the alternative scenario. If an interaction between fetal growth and a common factor is responsible for the observed association, it would be very grueling to uncover the true responsible interactive and perhaps nonlinear factor.

CONCLUSIONS

After considering the assumptions of Basso et al., several questions merit deliberation to determine the future course of research. First, is the assumption of no causal link between birth weight and neonatal mortality correct? If so, Basso et al. have shown us that the shape of the birth-weight-specific mortality curve might result from the presence of very rare confounders with very large effects. This finding suggests a search for the proverbial needle in the haystack. Second, if a causal link does exist, what is the nature of this link? If we relax the model to allow a causal link, the inverse J pattern of birth-weight-specific mortality could be also explained by a range of more common and weaker confounders that might have nonlinear effects and might interact with other risk factors. That is, speculations under the relaxed assumption basically take us to square one, where, if there is a needle in the haystack, we would have to accidentally sit on it to find it. Interestingly, perhaps we have to conclude that unrestricted models are correct, but often useless.

We have shown how the model used by Basso et al. may fail because of violations of the assumptions on which it is based. However, we should heed the lessons of their simple model, which opens up the appealing possibility of no direct causal effect between birth weight and neonatal mortality and shows how, in this context, the putative confounder would have to be rare and extremely strong. From a methodological point of view, their findings should change the way we consider birth-weight data when evaluating the effect of other risk factors on perinatal outcomes (2). More importantly, if they are right and no causal link exists, our research efforts to reduce neonatal mortality would be well advised to shift away from birth weight and toward direct causes of infant mortality and morbidity and toward broadening our efforts to acknowledge this possibility.

ACKNOWLEDGMENTS

Supported by the Intramural Research Program of the National Institutes of Health, National Institute of Child Health and Human Development.

Conflict of interest: none declared.

REFERENCES