evidence did finally, around 1976, result in the launch of such campaigns.

Conflict of interest: None declared.

References

Commentary: Resolutions of the birthweight paradox: competing explanations and analytical insights

Tyler J VanderWeele
Departments of Epidemiology and Biostatistics, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115, USA. E-mail: tvanderw@hsph.harvard.edu

The analyses in Yerushalmy’s paper1 indicated that, among low birthweight infants of less than 2500 g, maternal smoking was associated with lower infant mortality. The results have been replicated in a number of studies and populations, and these seemingly paradoxical associations are now often referred to as the ‘birthweight paradox.’ As can be seen from the present set of commentaries, Yerushalmy’s paper continues to generate discussion and interest even 40 years after its publication. Several explanations have been put forward for this paradoxical association. Here we will review some of these, discuss some relations and differences between the explanations that have been proposed, and then conclude with the question of what, if anything, we can infer about the role of birthweight in governing the associations between maternal smoking and infant mortality.

Explanations of the birthweight paradox

One explanation for the birthweight paradox associations is selection or confounding bias due to conditioning on an intermediate.2–4 If, as in Figure 1, there are unmeasured common causes (U) of low birthweight (L) and infant mortality (Y), such as for instance malnutrition or birth defects, then analyses of the association between maternal smoking (S) and infant mortality among low birthweight infants (L=1) can be in the reverse direction compared with that which one would have obtained if adjustment for U had been possible.

The intuition behind this explanation is that low birthweight might be due to a number of causes: one of these might be maternal smoking, another might be instances of malnutrition or a birth defect. If we consider the low birthweight infants whose mothers smoke, then it is likely that smoking is the cause of low birthweight. If we consider the low birthweight infants whose mothers do not smoke, then we know maternal smoking is ruled out as a cause for low birthweight, so that there must have been some other cause, possibly something such as malnutrition or a birth defect, the consequences of which for infant mortality are much worse.2 By not controlling for the common causes (U) of low birthweight and infant mortality, we are essentially setting up an unfair comparison between the smoking and non-smoking mothers. If we could control for such common causes, the paradoxical associations might go away.

Importantly, we can observe such paradoxical associations due to unmeasured common causes, even if there is no actual effect of low birthweight on infant mortality or even if there were no effect of maternal smoking on infant mortality at all. The association is induced by conditioning on the variable low birthweight L. The phenomenon is sometimes referred to, in the causal graphical models literature, as ‘collider stratification’ or of conditioning on a ‘collider’2–5 (i.e. a variable which is a common effect of
two other variables, S and U in this case, which can induce association between S and Y, even if there is no S->Y arrow, due to the path S->L<U->Y which generates association between S and Y because of conditioning on the collider L). Whereas it has not been definitively established that this is in fact the explanation of the birthweight paradox, it is now clear that very reasonable values for the effects of such unmeasured common causes U on low birthweight and on infant mortality would suffice to give paradoxical associations of the magnitude actually found in the data.\textsuperscript{3,4}

Another explanation that has been put forward for the birthweight paradox has been heterogeneity of the low birthweight phenotype.\textsuperscript{6} It may be that low birthweight infants, in fact, form two distinct subgroups. Some of these low birthweight infants may be reasonably healthy and just happen to fall in the lower tail of the normal distribution of birthweight. These low birthweight infant may not have poor mortality outcomes. Wilcox\textsuperscript{6} refers to such low birthweight infants as falling in the tail of the ‘predominant’ distribution. The other subgroup of low birthweight infants are those born prematurely and whose development may be problematic. Wilcox\textsuperscript{6} calls this the ‘residual’ distribution of low birthweight. Evidence for this distinction comes from the fact that if one attempts to fit a normal distribution to birthweight data, about 2–5% of births are in the lower tail in excess of what would be suggested by the normal distribution.\textsuperscript{6}

If we do have two distinct low birthweight phenotypes, then the birthweight paradox may be generated by neglecting this distinction. Among the non-smoking mothers, many of the low birthweight infants are likely from the ‘residual’ distribution and have poorer infant mortality outcomes. For the smoking mothers, maternal smoking in fact shifts the mean of the ‘predominant’ distribution lower. Thus, the low birthweight infants, those less than 2500 g, will include the ‘residual’ distribution but also a larger proportion of the ‘predominant’ distribution. When low birthweight infants of smoking and non-smoking mothers are compared, it might look as though the mortality outcomes of infants of smoking mothers are better because the low birthweight infants of non-smoking mothers are mostly those from the ‘residual’ distribution, whereas those of the smoking mothers are more of a mix of the ‘residual’ and ‘predominant’ distributions. But this might simply be an artefact of smoking shifting the mean of the ‘predominant’ distribution downward. Evidence that this may be what is taking place comes from comparing weight-specific mortality curves among smoking and non-smoking mothers, when using absolute birthweight vs z-scores for relative birthweight (calculated separately for smoking and non-smoking mothers to try to get at the differences in the predominant distribution).\textsuperscript{6} When comparing weight-specific mortality curves among smoking and non-smoking mothers using absolute birthweight, it looks as though infants of smoking mothers have better mortality outcomes at low birthweights. When comparing weight-specific mortality curves among smoking and non-smoking mothers using z-scores for birthweight, the mortality curve for infants of smoking mothers is higher than for those of non-smoking mothers at all birthweights.

In fact, the structure of the bias for this multiple low birthweight phenotypes explanation is not so very different from that of the unmeasured confounding explanation. Consider Figure 2a, wherein L1 indicates low birthweight from the predominant distribution and L2 indicates low birthweight from the residual distribution. Note that in this diagram we have, for the time being, no unmeasured common causes of the outcome and the low birthweight phenotypes. If in the analysis we collapse over these two distinct low birthweight phenotypes as in Figure 2b, with L again indicating any low birthweight, we can once again see from this diagram that S and Y would be associated conditional on L, through the path S->L<->L1<->L2<->Y. This association would still be present even if smoking had no effect on infant mortality whatsoever (i.e if both the S->Y arrow and the L1->Y arrow were absent). The structure of the bias is similar to that in Figure 1 insofar as we are conditioning on a common effect L, of both S and a variable associated with Y, namely L2.\textsuperscript{5} In Figures 2a and 2b, although low birthweight phenotype heterogeneity is responsible for the bias, rather than unmeasured confounding, the structures of the biases are quite similar. The structures would likewise be similar if the ‘residual’ low birthweight phenotype L2 did not directly affect infant
mortality (Y), but if instead L2 and Y had a common cause U as in Figure 2c. Once again conditioning on the collapsed low birthweight variable L would induce association between smoking and infant mortality even if there was no effect of smoking on infant mortality.

Other approaches for avoiding the birthweight paradox have also been proposed. It has been pointed out that if what one is interested in is simply the effect of maternal smoking on infant mortality then there is no need to adjust for birthweight at all.7 Birthweight may or may not be on the pathway from maternal smoking to infant mortality but it is almost certainly not a confounder.7 Thus, if what is of interest is the effect of maternal smoking on infant mortality (rather than say the direct effect of maternal smoking not through birthweight)7 then there is no need to adjust for low birthweight. When adjustment is not made for low birthweight, the birthweight paradox is avoided. Others have advocated a fetus-at-risk approach which focuses attention on effects at different gestational ages and uses as a denominator for risk, not all births, but rather only those fetuses that have survived up to a specific gestational age. In proceeding in this manner, the birthweight paradox once again vanishes.8,9 These insights may in fact be the most important points that have emerged from the discussion of the birthweight paradox. However, they do not really provide an explanation for the paradoxical associations between maternal smoking and infant mortality when conditioning on low birthweight; they merely provide a way of avoiding the paradoxes to begin with.

Empirical and structural distinctions

Our focus here is and has been on explanations of the paradoxical associations themselves as presented by Yerushalmy.1 The explanations given above concerning confounding/seletion on the one hand, or multiple low birthweight phenotypes on the other, both seem as though they might be plausible. To what extent can they be distinguished? The two explanations do have somewhat different empirical implications. Under Figure 1, if we were able to control for the common causes (U) of low birthweight and infant mortality, then the biases should vanish; an estimate of the association between maternal smoking and infant mortality among low birthweight infants controlling for U (e.g. malnutrition/birth defects) should indicate either a causative or null association, rather than a protective association. Under Figure 1, if we have controlled for common causes of low birthweight and infant mortality, we can then get valid estimates of the ‘controlled direct effects’10–12 of maternal smoking on infant mortality not through low birthweight, and these effects should be null or causative if there is indeed no actual protective effect of maternal smoking. Under Figure 2, if we could control for both the ‘predominant’ low birthweight phenotype (L1) and the ‘residual’ low birthweight phenotype (L2) as separate variables, rather than the collapsed variable L, then the biases should once again vanish; an estimate of the association between maternal smoking and infant mortality among low birthweight infants controlling for L1 and L2 should indicate a null or causative association, rather than a protective association. Under Figure 2, if we have controlled for both low birthweight phenotypes (L1 and L2), we can then get valid estimates of the ‘controlled direct effects’10–12 of maternal smoking on infant mortality not through low birthweight. In trying to distinguish between these two explanations, it might be of interest to examine empirically with data whether it is in fact possible to dissolve the birthweight paradox by controlling for such as malnutrition and birth defects on the one hand (the unmeasured confounding explanation) or by controlling for both the ‘predominant’ and ‘residual’ low birthweight types on the other (the multiple low birthweight phenotypes explanation). Obtaining adequate data
for such control may be challenging but is perhaps not inconceivable.

Of course, the reality may be a more complex interplay between these two explanations. The variable $U$ in Figure 2c might for example also affect $L_1$, in which case both unmeasured confounding and multiple low birthweight phenotypes could induce bias. Alternatively there might be a separate unmeasured common cause $U^*$ of $L_1$ and $Y$, and once again both unmeasured confounding and multiple low birthweight phenotypes could be involved in the explanation. Figures 1 and 2a-c in some sense represent idealized cases wherein only one of the two explanations is operative, but the reality may be that some more complex combination of these two, or other, explanations is in fact in play.

Even the two idealized cases in Figure 1 and Figures 2a–c are, in fact, not so far from each other. The structure of bias in Figure 1 that arises due to uncontrolled confounding by $U$ could potentially also be understood as outcome heterogeneity. Suppose we abandon the proposed distinction between ‘predominant’ low birthweight and ‘residual’ low birthweight, but that something like malnutrition or a birth defect serves as a common cause of low birthweight phenotypes could be involved in the explanation. Figures 1 and 2a–c in some sense represent idealized cases wherein only one of the two explanations is operative, but the reality may be that some more complex combination of these two, or other, explanations is in fact in play.

The role of birthweight

We have discussed possible explanations of the birthweight paradox associations. However, we are still left with the question as to what the role of birthweight is in the association between maternal smoking and infant mortality. Is low-birthweight in some sense responsible? In his paper Yerushalmy, in the discussion of the ‘displacement hypothesis’, seems to suggest ‘no’.1 In Table 8 of his paper he poses a counterfactual query of the form of what would have happened had the smoking mothers not smoked. He notes that there would have been fewer low birthweight infants among this group of actual smokers but, because infant mortality is higher for low birthweight infants of non-smoking mothers, he concludes that the overall net effect if smoking mothers had not smoked is negligible. However, in his analysis, he relies on the infant mortality rates for the non-smoking mothers in addressing his counterfactual query. In other words, he ignores the issue that the low birthweight infants among the smoking and non-smoking mothers may be different (that those of the non-smoking mothers may have fewer ‘predominant’ low birthweight infants, or may have fewer malnutrition-caused instances of low birthweight). We might instead then ask what would have happened to infant mortality had the smoking mothers not smoked and this altered their low birthweight distribution but if we kept the infant mortality rates among the smokers in calculating the final infant mortality rate for this group? A counterfactual query of this form is in fact precisely what the newly developed methods for mediation analysis address.10–14 The natural indirect effect, which is the target of these methods, asks: What would have happened to infant mortality among smokers if we could block the effect of smoking on low birthweight so that the smoking mothers had the same rates of low birthweight as the non-smoking mothers?

If we apply such methods12–14 using 2003 US birth certificate infant mortality files from the National Center for Health Statistics with: maternal smoking as exposure; low birthweight (less than 2500 g) as the mediator; and infant mortality as the outcome; while controlling for maternal drinking, age, race, marital status and education; then the overall effect of maternal smoking is to increase the likelihood of infant mortality by an odds ratio of 1.48 [95% confidence interval (CI): 1.41, 1.54] and the natural indirect effect odds ratio12,14 estimate is 1.31 (95% CI: 1.30, 1.33). In other words, the effect of smoking on infant mortality simply by increasing low birthweight among smokers is to increase the likelihood of infant mortality by an odds ratio of 1.31. The proportion of the effect of maternal smoking on infant mortality mediated, on a risk difference scale,14 is estimated to be 74% (95% CI: 67%, 81%). Low birthweight at least on the surface in fact appears to play an important role.

We might of course be concerned about unmeasured confounding as in Figures 1 or 2c. However, we can use sensitivity analysis techniques to assess how such unmeasured confounding might change estimates.15 If an unmeasured confounder $U$ had: a prevalence of 2% among the smoking and non-smoking mothers with normal birthweight infants; a prevalence of 3% among smoking mothers with low birthweight infants; and a prevalence of 9% among non-smoking mothers with low birthweight infants (i.e. if for those with low birthweight infants $U$ was three times more likely among non-smokers than smokers); and if $U$ increased the likelihood of infant mortality by 3-fold; then the proportion mediated would decrease to 64%
the effect of maternal smoking on infant mortality is mediated by unmeasured confounding. But unmeasured confounding does not seem to substantially alter the conclusion that a considerable portion of the birthweight paradox is present with these data as it was in Yerushalmy’s original paper. With the 2003 data, the odds ratio for infant mortality comparing smoking and non-smoking mothers with low birthweight infants is 0.93 (95% CI: 0.88, 0.97), seemingly suggesting a protective effect of maternal smoking. Interestingly, although the conclusion that low birthweight mediates a substantial portion of the effect of maternal smoking on infant mortality is fairly robust to unmeasured confounding, the birthweight paradox itself is more easily explained away by such unmeasured confounding. Again using sensitivity analysis, even in the more modest unmeasured confounding scenario considered above, with 3-fold increases in the risk, the odds ratio for infant mortality comparing smoking and non-smoking mothers with low birthweight infants shifts to 1.03 (95% CI: 0.98, 1.08) and the birthweight paradox thus vanishes. In summary, unmeasured confounding may be responsible, at least partially, for the birthweight paradox, but unmeasured confounding does not seem to substantially alter the conclusion that a considerable portion of the effect of maternal smoking on infant mortality is mediated by low birthweight.

We might, however, also be concerned about the interpretation of these effect estimates if low birthweight does not in fact itself cause infant mortality but is simply serving as a proxy for something else—the developmental process or prematurity—that does. Recent theory in the mediation analysis literature indicates that the indirect effect estimates can still be interpreted as the portion of the effect mediated by the processes (e.g. prematurity) leading to low birthweight that are captured by the low birthweight measurement. The low birthweight measurement does not of course capture all aspects of the underlying processes and so it is possible that our assessments of the proportion mediated are underestimates. However, regardless, it does seem that we have fairly substantial evidence that either low birthweight, or perhaps more plausibly the developmental processes leading to it, are important in understanding the association between maternal smoking and infant mortality.

Conclusion

Where are we then in our understanding of the birthweight paradox? Do we have explanations for why these paradoxical associations exist? Yes, at least two: unmeasured intermediate-outcome confounding and multiple low birthweight phenotypes. Do we have ways to avoid the birthweight paradox? Yes, by not conditioning on birthweight at all (which will often be the correct analysis) or by using a fetuses-at-risk approach. Is low birthweight important in explaining the effect of maternal smoking on infant mortality? Yes, low birthweight or the processes that lead to low birthweight are relatively important. Do we know whether it is unmeasured confounding, or multiple low birthweight phenotypes or some other explanation that is in fact primarily responsible for the birthweight paradox? No, this is still an open question and it could well be some complex interplay between the various explanations that have been put forward.

Funding

Tyler VanderWeele was supported by NIH grant R01 ES017876.

Conflict of interest: None declared.

References

Commentary: Smoking, birthweight and mortality: Jacob Yerushalmy on self-selection and the pitfalls of causal inference

Mark Parascandola
Tobacco Control Research Branch, National Cancer Institute, 9609 Medical Center Drive, Bethesda, MD 20892-9761, USA. E-mail: paramark@mail.nih.gov

In the television show Mad Men, which goes to great lengths to maintain historical accuracy as to its 1960s setting, a visibly pregnant Betty Draper is shown smoking in the maternity ward. Little research had been conducted at the time on the potential effects of smoking during pregnancy. Jacob Yerushalmy’s 1971 paper appeared as a wave of new findings pointed towards the detrimental impact of maternal smoking on the developing fetus. But the paper goes beyond merely reporting study results. Yerushalmy methodically sifts through the data from multiple angles, testing a range of potential hypotheses, while constructing an argument about the perils of causal inference from observational studies. The paper was one of a series, part of a discussion that persisted from the 1950s to the 1970s and possibly beyond. Understanding this historical context is essential to appreciating Yerushalmy’s contributions.

On epidemiological methods and causal inference

Yerushalmy (1904–73) was born and raised in what is today known as Israel, until he came to the USA in 1924 to attend university, studying mathematics at Johns Hopkins University. Later, as an instructor at Hopkins, home to Raymond Pearl and Wade Hampton Frost, he was exposed to the growing field of biostatistics. Yerushalmy held a series of posts as a statistician in the New York State Department of Health, the Public Health Service and the Children’s Bureau of the Department of Labor, where he began studying child development. In 1947 Yerushalmy moved to the University of California at Berkeley to found a new department of biostatistics. Beginning in 1959, he also led the Child Health and Development Studies (CHDS), a cooperative project with Kaiser Permanente Hospital. The CHDS was innovative in conducting long-term follow-up of participants in one of the first health maintenance organizations. This ongoing prospective study became the locus of Yerushalmy’s work throughout the following decade.

During the 1950s, Yerushalmy became an active commentator on epidemiological methods and causal inference, bringing attention to sources of bias and flawed analyses. For example, along with Herman E Hilleboe of the New York State Department of Health, he criticized ecological studies comparing dietary fat and heart disease mortality across a selection of countries. They pointed out that the countries differed in many respects beyond the two variables of interest, and also that the seemingly arbitrary selection of countries for analysis tended to favour the result. Investigators should, they recommended, assess whether ‘the association between two variables is in fact between the variables investigated and does not merely reflect relationships with a broader group, of which one or the other of the variables forms a part’. But most notable was the influential 1959 paper by Yerushalmy and Palmer, which urged that epidemiologists should seek to imitate ‘the more rigorous methods long in use by bacteriologists’. Bacteriologists had a set of rules—Koch’s postulates—for drawing aetiological conclusions about infectious agents. But these methods relied on identification of specific disease agents and laboratory