Invited Commentary

Invited Commentary: Fewell and Colleagues—Fuel for Debate

James Marshall

From the Division of Cancer Prevention and Population Sciences, Roswell Park Cancer Institute, Buffalo, NY.

Received for publication March 31, 2006; accepted for publication July 12, 2006.

Concern over the impact of flawed measurement continues to nag epidemiology. Early studies indicated that the impact of measurement error is benign, leading generally only to attenuation of associations; more recent research has documented that this impact, especially within the setting of multivariate modeling, cannot be expected always to be benign. It can, for example, be a source of unsettling inconsistency. Fewell and colleagues (Am J Epidemiol 2007;166:646–655) show that residual confounding is especially persistent in the presence of multivariate confounding.

Extending our understanding of multiple, imperfectly measured confounders, Fewell et al. (1) have documented that confounding is not eliminated by adjustment for a mélange of possible confounding factors. They have shown that error in the measurement of multiple confounders will cause them to resonate through the focal epidemiologic exposures and bias estimates of their effects. These results confirm that, absent understanding of the extent and structure of measurement error in our study variables and their confounders, we need to very carefully circumscribe our claims that our multivariate analyses eliminate the threat of confounding.

Fewell et al. confirm that confounder misclassification complicates the interpretation of epidemiologic data. Epidemiology has long held to Bross’ (2) and Newell’s (3) demonstrations that completely random misclassification of a binary exposure reduces test power and biases estimated associations toward the null; their results seemed reassuring insofar as one might be most concerned about false-positive findings, although Goldberg (4) later demonstrated that modest departures from random behavior of measurement errors could generate bias away from the null. All of these investigations came before epidemiologists worried much about multivariate analysis. That exposure-misclassification results do not apply in the case of a mismeasured confounder was demonstrated by Greenland (5) and Greenland and Robins (6). More recently, Fox et al. (7) illustrated the importance of seemingly trivial departures from random behavior of measurement error.

Exploring the problem of multiple confounders in epidemiology, Fewell et al. (1) raise the specter of numerous troublesome scenarios: Several confounders could be measured with varying degrees of error; the degree to which confounders actually confound could vary; combinations of these conditions could be present. In the examples Fewell et al. provide, all of the confounders increase risk, and all are positively related to one another; the result will certainly be different if—and this is possible—some of these positively correlated confounders increase risk while others decrease it. In these examples, the confounders are truly confounders: If they are correlated with risk and with the focal exposure but do not actually have a direct bearing on risk, controlling for them may induce bias. In these examples, all measurement errors are well-behaved: Independently of risk and of one another, they may well be correlated.

One never really knows what process gave rise to data-based results. The most significant limitation of the results Fewell et al. develop is that the focal exposure is measured without error. It is, of course, possible that measures of one exposure will contain no error while measures of other exposures will contain substantial error; whether this is likely

Correspondence to Dr. James Marshall, Division of Cancer Prevention and Population Sciences, Roswell Park Cancer Institute, Elm and Carlton Streets, Buffalo, NY 14263 (e-mail: james.marshall@roswellpark.org).
might be debated, and it would be difficult for the investigator to justify counting on it. The absence of error in measurement of the focal exposure eliminates the possibility of correlations between that exposure’s measurement errors and those of the confounders. Such correlations can have significant effects on what is observed, and those effects are not readily derivable from what is seen in the case of well-behaved error (8).

Fewell et al. qualify their results, noting that those results are based on simulation (1). Asserting that results based on data are always desirable, they recognize that such results may not be available. However, it is worth noting that one never really knows what process generated observational data. The distinction between a set of observations and the process that gave rise to those observations is important. We know and can readily understand the process that produced the data Fewell et al. address.

Fewell et al. constructed a set of confounders correlated at either 0.1, 0.3, or 0.5 with the focal exposure (1). However, if the true correlations of the confounders with the focal exposure are 0.1, 0.3, and 0.5, respectively, the observed correlations of these variables with that exposure will not be 0.1, 0.3, or 0.5. With the intraclass correlation coefficient of each confounder being equal to 0.5, the observed correlations will be 0.07, 0.21, or 0.35. With the exposure measured with the same intraclass correlation coefficient as the confounders, the observed correlations will be 0.05, 0.15, or 0.25 (9). Measurement error substantially attenuates these correlations, suppressing evidence of the degree to which the confounders confound.

Recognizing the limitation imposed by their modeling of no exposure measurement error, Fewell et al. suggested that it would be valuable to study the impact of error in measurement of the exposure as opposed to that of the confounder (1). Some limited data on this problem have been generated; in general, the bias imposed by a mismeasured confounder depends on whether the confounder increases or decreases risk, as well as on whether the confounder is positively or negatively associated with the focal exposure (9). In more complicated settings, such as those with correlated measurement error, these effects can be more varied.

These results underline the persistence of bias in spite of control for additional mismeasured risk factors. The problem is especially acute for multiple orthogonal variables. In Fewell et al.’s table 3 (1), with four imperfectly measured but uncorrelated confounders, for example, the correlations of $E$ and $X_1$ and $E$ and $X_2$ are 0.3; the intraclass correlation coefficients for $X_1$ and $X_2$ are 0.5. The crude odds ratio describing the impact of the exposure is 2.35; the odds ratio adjusted for $Z_1$ is 2.24, and that adjusted for both $Z_1$ and $Z_2$ is 2.12. Control for two important variables barely reduces bias; most of it is preserved. Fewell et al.’s table 4 (1) shows that the problem remains even for correlated variables. In the same setting with correlated variables, as seen in Fewell et al.’s table 4, the crude odds ratio is 1.89, while the odds ratio controlled for both $Z_1$ and $Z_2$ is 1.58; again, most of the confounding imposed by measurement error is preserved.

Fewell et al. state that their results apply most specifically to the situation in which exposure does not affect risk (1). Of course, the interpretation of any apparent risk alteration—miniscule or substantial—requires understanding of the extent and structure of errors underlying exposure and confounder measurement. One can neither establish nor dismiss the importance of exposure to risk, absent information on the degree and structure of measurement error. One can readily construct from a data set—an outcome and series of exposures—using nothing more than measurement error, an infinity of apparent impacts linked to those exposures.

This thoughtful analysis confirms that measurement errors bias the expected value of association measures. In other words, notwithstanding sample size, notwithstanding statistical significance, measurement error is more than a benign influence that leads us to occasionally overlook risk factors. What we are increasingly having to recognize is that measurement error can lead, under a broad array of circumstances, to substantial bias. We cannot abstract from our data the correct conclusions unless we thoroughly understand what we have measured and how well we have measured it. Passing glances at the repeatability of one or two variables out of a series of variables, even if reinforced by claims that these analyses indicate that the data are valid, will not do. Sanguine reassurances that the data are probably “good enough” (10, 11) are likely to contribute little more than false confidence in our results. As we seek to refine epidemiology’s contribution, we may need to devote more attention than we have heretofore to the quality of our data, assaying that quality and allocating resources to increase it.

It is, of course, one thing to recognize measurement error as an obstacle to the interpretation of epidemiologic findings and quite another to do something about it. One of the most comprehensive approaches has been provided by Spiegelman et al. (12, 13). A more modest approach, proposed as a ready guide to the interpretation of categorical data, was described by Fox et al. (7).

We need to understand how measurement errors are correlated. If we assume incorrectly either that these errors are uncorrelated or that they are correlated, we run the risk of extracting the wrong conclusions from our data.

It may also be worth remembering, as we struggle with the formidable challenge of examining factors that cannot readily be measured other than very crudely, the importance of triangulation: juxtaposing epidemiologic data against in vitro and in vivo findings, against human experimental findings, and even against ecologic evidence (14). To be sure, probability theory tells us that individual-based analysis will give us the right answer over the long run, if we specify our epidemiologic models correctly, measure our variables accurately, and draw our samples appropriately. We are never sure we have the correct model, are virtually never sure we have measured our variables adequately, and are almost never sure we have drawn our samples appropriately; and significant violations of any of these conditions may mean that bets are off. These biasing effects may tend to cancel one another, and we do not at present have enough information to justify relying on such benign cancellations. In that our job is to help people understand the impact of their experiences and exposures on the probability that they will get sick, not to preserve our approach to information collection, we must recognize the limits of and remain open to all of the information to which we have access.
ACKNOWLEDGMENTS

The author thanks Dr. Sander Greenland for helpful advice.
Conflict of interest: none declared.

REFERENCES