Invited Commentary: Shifting the Burden of Proof Regarding Biases and Low-Magnitude Associations

Irva Hertz-Picciotto

The issue of what exactly are the limits of epidemiology is now on the radar screen of the entire field. Shapiro (1) has added his thoughts to the public discussion by describing a sensitivity analysis of reported associations between oral contraceptive use and breast cancer. He addresses diagnostic selection bias and information bias as possible explanations for small relative risks, demonstrating quite clearly that the overall association of 1.07 for ever use of oral contraceptives reported in a “collaborative reanalysis” of 54 studies could easily be due to a rather low magnitude of each of these biases.

His points are reasonable, but limited; his overall conclusions, however, are a mixed bag. On the one hand, he correctly cites the need to separate policy decisions from the strength of evidence regarding causality. I will return to this question at the end. On the other hand, he essentially advises that if bias can explain small associations, we should give up our scientific pursuit of the health issue in question (“...it is unlikely that refinements of epidemiologic or statistical methods, sensitivity analyses, validations studies, etc., can get around the difficulty” (1, p. 944)). This conclusion is not scientifically grounded for reasons I will discuss.

1) If bias is to be put forward as the hypothesized explanation for a set of findings, it should be subject to the same scrutiny as the hypothesis of causality: Shapiro asserts that, compared with women who do not take oral contraceptives, those who do are more likely to be diagnosed with breast cancer or are likely to be diagnosed at an earlier stage. Undoubtedly, there are diseases for which this diagnostic bias is likely. However, what evidence is there that gynecologists are more likely to instruct patients on breast self-examination or to perform such examinations themselves when prescribing oral contraceptives compared with either the same physicians seeing patients for other reasons (for other forms of contraception or for gynecologic problems) or with nongynecologists? Have studies of mammography utilization demonstrated higher use by women taking oral contraceptives? If there is a “lead-time” bias, are the differences between exposed and unexposed with respect to the time to diagnosis of sufficient magnitude? Shapiro’s calculations (1, table 4) are all predicated on an average 6-month advancement in diagnosis.

For former oral contraceptive users, the plausibility of diagnostic bias seems more remote; speaking from personal experience, I have never had an allopathic (Western) physician ask me about my prior use of oral contraceptives. Nor does the American Cancer Society or any other organization recommend a more intensive schedule of mammography for women who currently use or formerly used oral contraceptives. Since the largest group of women at risk for breast cancer is beyond reproductive age, there would appear to be no basis to conclude that screening of former oral contraceptive users is more intense than is screening of never users. While Shapiro’s hypothesis is not impossible—women taking oral contraceptives may be more frequent users of health care generally, regardless of the actual advice they receive—evidence should be garnered to examine it. (Interestingly, the published reanalysis that Shapiro critiques (2) actually includes data on the use of mammography among controls; in comparison with those who never took oral contraceptives, current and recent users were less likely (not significantly) to have had a mammogram, while users more than 10 years previously were more likely to have had a mammogram (2, appendix 54). Given the stronger association with current or recent use, these data demonstrate the reverse of what Shapiro proposes.)

2) For a sensitivity analysis to shed light on the plausibility of a causal versus an artifactual association, it should attempt some degree of comprehensiveness: Shapiro is convinced that selection and information biases in studies of oral contraceptives would both weigh in the direction of exaggerating risks. His fail-
to consider a major source of bias in the opposite direction is problematic. Considering the extent to which the hypothesis under study is well-known (as Shapiro points out, "...there was widespread and well-publicized concern about breast cancer risk from the time oral contraceptives were introduced") (1, p. 943), women with a family history of breast cancer may have either chosen to or been counseled by health professionals to avoid the use or prolonged use of oral contraceptives. Whether or not their higher family incidence is due to genetic or environmental factors is irrelevant; those at highest risk might well have been least likely to be exposed to oral contraceptives. The result would be downwardly biased associations. Again, the presence of this bias and its magnitude could be addressed by epidemiologic investigation. The implication for Shapiro's sensitivity analysis is that the "null" situation (no association between oral contraceptives and breast cancer) would give rise to a relative risk lower than 1.0; thus, if high-risk women have opted not to take oral contraceptives, a much greater diagnostic surveillance bias (how much greater? 10 percent? 20 percent?) might be needed to yield the observed relative risk of 1.07.

3) The approach to causal inference must draw on much more than the magnitude of an association: Scientific thinking proceeds by weighing the totality of evidence and considering the full range of alternative explanations in light of this body of data. Consider the oral contraceptive versus breast cancer evidence. Quite simply, what makes the data so unconvincing is that besides the very low odds ratios, there was no effect of duration/dose (2, 3). The clear inconsistency (with a causal interpretation) is that those who took oral contraceptives much longer ago had much higher doses, yet it is in those who took them most recently that the strongest association is observed. Hence, there appears to be no evidence of either a cumulative or a dose-rate effect. (Note that there are conditions when the lack of a dose response does not detract from a causal inference, viz., when higher doses cause different diseases that are incompatible with survival or that mask the disease under study; these conditions do not apply here.) In contrast to Wynder's admonitions to take into account the full range of evidence that bears on the inference of causality (4), Shapiro incorrectly bases his judgment on the small association alone.

The narrow focus on the magnitude of the association leads Shapiro astray when he tries to generalize to other "small associations," including that between passive cigarette smoke and lung cancer. Here, a wide array of observations provide a rather convincing picture: 1) mainstream cigarette smoke causes lung cancer, and sidestream smoke represents a lower dose of a similar chemical mixture; 2) sidestream smoke actually includes higher concentrations of some very potent carcinogens compared with mainstream smoke (5); 3) linear extrapolation from the active smoker's dose to the lower passive smoker's dose predicts risks similar to the magnitude of those observed; and 4) these low doses are of clear biological/clinical relevance, given evidence of rather strong associations with sudden infant death (6) and respiratory problems in children exposed to second-hand smoke (7, 8). Thus, the evidence supports biologic plausibility, coherence, and dose response. Even so, the appropriately skeptical epidemiologist must ask the questions that Shapiro (1) is raising: What about information or selection bias? More concretely, what evidence is there for these biases? Some studies have shown that spouses underreport the vices of their mates, but do cases of lung cancer underreport partner's smoking less than non-cases? Do spouses of smokers undergo better surveillance than spouses of nonsmokers?

In light of the above and, indeed, for any small or even moderate associations, I suggest that the question is not whether we can devise some scenario that could produce the relative risks that were observed. Rather, we should ask: 1) what evidence exists that upward biases are present and that they outweigh biases in the other (downward) direction? and 2) how does the evidence for such biases compare with the evidence for causality, i.e., is it more or less convincing?

4) The patterns of systematic biases are not uniform across the field of small associations: Shapiro's broader generalizations imply near-universality of the problems on which he has focused. For some research questions, exposures can be ascertained from objective sources, reducing or eliminating differential information bias. Similarly, high participation rates can be achieved; the study question does not need to be, and in many epidemiologic studies is not, revealed to participants; and follow-up can be relatively complete. In some instances, downward biases may outweigh upward ones (as described above, this scenario may be the case for Shapiro's example). For environmental exposures, selection bias might tend to remove more exposed than unexposed cases due to their lower socioeconomic status (environmental exposures tend to concentrate in low socioeconomic neighborhoods) and the associated reduced access to medical care; this selection differential (by exposure) would tend to be greater for cases than for controls when the latter are population based, thereby creating the conditions for downward bias.

Before concluding, I must commend Shapiro for two of his remarks, one minor and one major. The minor one is his injunction that the reporting of odds ratios to
the third significant digit (for positive associations, two digits after the decimal point) is pseudoprecision. Actually, in the example he critiques, it is "pseudoaccuracy," not "pseudoprecision," since a very large study does give very high precision but no guarantee of accuracy, i.e., lack of bias. However, for most studies published in epidemiologic journals, it is both, unfortunately. His second, more significant comment concerns the clear separation between where science ends and where policy begins. He suggests that it might be prudent policy, in light of the uncertainty of the knowledge base and the possibility that the absolute numbers affected could be large, to "act as if such associations are indeed causal" (1, p. 944), while recognizing that such decisions are a matter of judgment and are outside the realm of science. Thus, it may be justified to inform the public, in an evenhanded manner as possible, that the evidence is not conclusive but that science has not been able to exclude a given risk. Such a statement leaves the individual or the group, not the scientist, to make the decision in the face of such uncertainty.

While I also support Shapiro's skepticism about small associations, I challenge his conclusions about the limits of epidemiology. In doing so, I am making a fundamental argument about the burden of proof. Just as any claim of causality should be evaluated to assess consistency with all of the available information, any claim of particular biases should undergo equal scrutiny. This principle applies to selection bias, measurement error (including information bias), and confounding. It is common practice for those who are criticizing a study to invoke the possibility of unmeasured confounders. Frequently, however, no specific variable is cited, or if it is, no argument or evidence is given for why it should be associated with both exposure and outcome and in what direction those associations would likely be. Speculation about biases, to be useful, should propose "testable" hypotheses (here, I do not refer to significance testing but to the accrual of data in favor of or counter to a theory). Validation studies can be conducted to determine the degree of information bias, and data on health care utilization can be used to evaluate the magnitude of diagnostic bias. In fact, recommendations to collect these types of data for studies of breast cancer and oral contraceptives were published in 1988 (9). In short, Shapiro's arguments (1) about the futility of disentangling small associations are unconvincing in the empirical example he provides, and their applicability beyond that situation is unclear.

It is time to raise the bar for those who believe they have found the limits of epidemiology. We need more, not fewer, sensitivity analyses, refinements in methods (such as corrections for measurement error), and validation studies. We need hard data to support or refute the potential role of biases. Those who are skeptical of alleged causal associations (including myself) and who submit that they are due to bias should meet the same standard of proof as is required for those who continue to pursue small associations. That is the scientific method, and that is the way to find our limits.

REFERENCES