Is There Is or Is There Ain't No Baby?: Dr. Shapiro Replies to Drs. Petitti and Greenland

Samuel Shapiro

I interpret Greenland's (1) assertion that I wish to ban meta-analysis as poetic license, and I thank him and Petitti (2) for their thought-provoking responses. To me, one of the most remarkable features of this symposium is that there is little disagreement among us concerning the synthetic meta-analysis (to use Greenland's expression) of nonexperimental data. Greenland and Petitti have not challenged any of the examples that I criticized (3), nor have they offered other examples that they consider valid. What that implies, by extension, is that we agree that the great bulk of the meta-analyses published to date are of questionable validity.

The large area of agreement considerably simplifies consideration of our disagreements. The responses raise three issues that are at the crux of the argument. First, Greenland makes a valuable distinction between synthetic meta-analyses whose purpose is to produce a single summary risk estimate, and those "aimed at testing criticisms of study results and identifying patterns or trends in [those] results." He considers synthetic meta-analysis to be misleading, except in those rare instances in which all studies are in agreement (and in that case, meta-analysis is superfluous). He even proposes that when study results are so heterogeneous as to require specification of a random-effects term, they are meaningless. On that matter, there is not a hairsbreadth of difference between us.

By contrast, Greenland considers the quantitative identification and description of how and why studies agree and disagree with each other to be the valid and worthwhile potential contribution to knowledge offered by meta-analysis. That is, if I interpret him correctly, meta-analysis can be used for the purpose of quantitative literature review. I have my doubts about that application as well, but if it were the only one, it is unlikely that we would be engaged in this symposium: by far the majority of the published meta-analyses have been of the synthetic type. Nevertheless, it is worth considering why we disagree with regard to meta-analysis as literature review. Because of the unquantifiable but important subtleties involved in any such review, I question whether quantitative methods can ever be as thoroughgoing, probing, and informative as qualitative methods.

The relative merits of the two approaches are well illustrated by the coffee-myocardial infarction controversy (4) to which Greenland refers to support his argument. His meta-analysis shows that the early case-control studies and cohort studies disagreed, with the consequence that the association was imputed to case-control biases; but that conclusion proved to be premature, because later cohort studies documented positive associations.

My reasoning, based on a qualitative review of the evidence, is as follows: If coffee increases the risk of infarction, the biologic mechanism might well be the acute adverse effects of caffeine on cardiovascular function (5-7); based on pharmacodynamic
n matter, there is not a hairs-breadth between us. Greenland considers the identification and quantification of nonrandom errors a valid and worthwhile contribution to knowledge analysis. That is, if I interpret, meta-analysis can be a useful tool for quantifying the results of individual studies. This was only true if the only analysis we would be engaged in was one in which the majority of the meta-analyses have been of the same type. Nevertheless, it is worth noting that we disagree with regard to the value of literature review. Because of the importance of literature review, I question the value of methods that can ever be more, probing, and informative.

One of the two approaches favored by the coffee-myocardial-artery controversy (4) is to support his argument. It is shows that the early case-control studies do not always conform to the case-control biologic assumptions. If we reject meta-analysis because of the absence of an association, we nevertheless continue to perform them. We may be more useful to take the view of the quantitative and qualitative perspectives that complete each other. In any event, however, the real concern is with the synthetic uses of meta-analysis. No matter which approach we agree on, the remainder of this response, in principle meta-analysis will be used as shorthand for synthetic meta-analysis.

The second issue raised in Greenland's and Petitti's responses is the argument that at least as it applies to small effects, meta-analysis can be more useful to take the view that the quantitative and qualitative perspectives complement each other. In any event, we nevertheless continue to perform them. If we reject meta-analysis because of the absence of an association, we nevertheless continue to perform them. We may be more useful to take the view of the quantitative and qualitative perspectives that complete each other. In any event, however, the real concern is with the synthetic uses of meta-analysis. No matter which approach we agree on, the remainder of this response, in principle, meta-analysis will be used as shorthand for synthetic meta-analysis.

That example is especially instructive because, in an earlier report, we had identified weak association (10), and under a causal hypothesis, we may have been wrong if the suspicion is that meta-analysis is most tempting. In some instances, there may be strong a priori grounds for believing that an effect, although small, may be real, and if real, of great public health importance (e.g., cancer risk in relation to water chlorination (8)). However tantalizing that possibility may be, it is not a justification for meta-analysis; rather, if the suspicion is well grounded, it is a justification for further research that improves on the work performed to date.

Again, Greenland provides a useful illustration of the issues by referring to one of our studies of alcohol and breast cancer (9). That example is especially instructive because, in an earlier report, we had identified weak association (10), and under a causal hypothesis, we may have been wrong if the suspicion is that meta-analysis is most tempting. In some instances, there may be strong a priori grounds for believing that an effect, although small, may be real, and if real, of great public health importance (e.g., cancer risk in relation to water chlorination (8)). However tantalizing that possibility may be, it is not a justification for meta-analysis; rather, if the suspicion is well grounded, it is a justification for further research that improves on the work performed to date.

For these reasons, we performed an additional independent study that made good on some of the defects (9). The result was null. That finding, together with a qualitative review of the literature (11), leaves me skeptical as to any possible causal relation. On the other hand, the two meta-analyses that I referred to (12, 13) reached the opposite conclusion. Readers interested in...
reaching their own judgment as to which of the two approaches is the more valid might wish to compare the meta-analyses with the review.

The third issue raised by the responses is the post hoc, ergo propter hoc argument that since the new fashion of meta-analysis is unlikely to wither on the vine, we had better make the best of it. Meta-analysis will become an established part of the academic curriculum; there will be a cornucopia of funding for grants; and government departments will continue to make public health decisions, often misguided ones, based on the results of meta-analyses. We can play a role in keeping everything within reasonable bounds. Petitti also argues that qualitative review is an academic backwater, whereas meta-analysis has a certain appeal, and that we should do whatever we can to ensure that the limitations are recognized and the techniques used responsibly. I disagree: Bad science, however politically correct it may be for the moment, should be discredited as bad science.

The main issues having been responded to, a few remaining matters also call for response. First, Greenland offers elegant arguments that the definition of what constitutes a weak effect should vary according to context; and of course he is right in arguing that a relative risk of well below 2.0, or even 1.5, for a low exposure level can sometimes be interpreted as causal if higher levels of exposure produce higher risks (as with smoking). Nevertheless, in situations in which the highest risk that can be identified for any exposure category is less than twofold, I think my definition, although less elegant, is not materially different from his.

Second, Greenland offers some comments about data pooling, that is, the aggregation of "raw data" (as opposed to the meta-analysis of published data) from a range of studies. I did not consider data pooling because it fell outside the limits specified for the symposium. I am glad that the matter has been raised, however, because Greenland points out some important defects in data pooling, and also because there are a great many more that still need to be pointed out. The fashion has been accepted uncritically, and another symposium is needed.

Third, Petitti refers briefly to the application of meta-analysis to randomized controlled trials. Again, that topic was outside the scope of the symposium, and it deserves extensive debate in its own right.

Fourth, Petitti agrees with my criticisms of published meta-analyses, but she nevertheless feels that I am throwing out the baby with the bathwater. She argues that a "judgment about [the] promise [of meta-analysis] should not be based on the early studies that used the method." If so, where are the later studies that fulfill that promise? On conceptual grounds, can we expect there will be such studies?

Finally, I believe we are now confronted by a major educational dilemma. In recent years, meta-analysis has been uncritically embraced by many as a panacea. The scale of that embrace is unprecedented. There is hardly a medical journal in which it has not been claimed at some point that a "meta-analysis has shown that A causes B," or words to that effect. It took years to gain acceptance for the idea that \( p < 0.05 \) does not by itself indicate causation; it will probably take at least as long to drive home the limited interpretability, if not the lack of interpretability, of meta-analysis. One reason for the difficulty is that we have not yet set our own house in order; there are still too many epidemiologists who are willing to equate data aggregation with truth.

REFERENCES
more that still need the fashion has been and another symposi-

briefly to the appli-
s to randomized con-
hat topic was outside
> own right.

es with my criticisms
y based on the early
"If so, where
pounds, can we expect
es are now confronted
ile. In recent
as been uncritically
aneuca. The scale
point that a “meta-
a, or it took years to gain
is that p < 0.05 does
bution; it will prob-
ing to drive home the
, if not the lack of
alysis. One rea-
that we have not yet
order: there are still
ists who are willing
ation with truth.

6. Dobmeyer DJ, Stine RA, Leier CV, et al. The
arhythmogenic effects of caffeine in human be-
7. Robertson D, Curatolo PW. The cardiovascular
effects of caffeine. In: Dews PB, ed. Caffeine: per-
spectives from recent research. New York,
ination, chlorination by-products and
case-control study of alcoholic beverage con-
sumption and breast cancer. Am J Epidemiol
cancer and alcoholic-beverage consumption.
11. Rosenberg L, Metzger LS, Palmer JR. Alcohol
consumption and risk of breast cancer: a review
of the epidemiologic evidence. Epidemiol Rev
1993;15:133-44.
A meta-analysis of alcohol consumption in rela-
tion to risk of breast cancer. JAMA 1988;260:
653-6.
13. Longnecker MP. Alcoholic beverage consump-
tion in relation to risk of breast cancer: meta-
analysis and review. Cancer Causes Control
1994;5:73-82.

Source: https://www.industrydocuments.ucsf.edu/docs/zzjx0000