Commentary: Prior specification of hypotheses: cause or just a correlate of informative studies?

David A Savitz

Published epidemiological studies are prone to spurious positive findings. This is not just an issue bearing on the discipline’s credibility to outsiders but a fundamental methodological concern. Epidemiologists must accept the challenge to improve research methods, publish findings regardless of their implica-
tions, and objectively appraise the validity of our results and those of our colleagues. Results are often dichotomized as ‘positive’ or ‘negative’, despite the loss of quantitative information resulting from this practice. The aetiology of false positive reports is surely multifactorial. Some of this falls to the media and the public for overinterpretation. Some results from the exuberance of investigators who advertise their most surprising, dramatic findings, despite the fact that results that run counter to the conventional wisdom are most likely to be erroneous. Human beings (not just epidemiologists) can become enamoured with their own achievements, lose objectivity, and seek the fame and fortune that result from startling discoveries. We need to improve the resolution of our methods and devote greater energy to helping to ensure appropriate use (or lack of use, in many cases) of our findings by policy makers and the public.

Swaen and colleagues1 have taken on the important goal of improving understanding of the aetiology of false positive studies, searching for causes based on past research that could be applied to future studies to help distinguish between true positive and false positive findings. Such identifiers would enable us to place a more appropriate level of confidence in study findings, discounting some and paying more attention to others. The authors deserve credit for attempting to bring some empirical, quantitative evidence to bear on this important issue, but some practical and conceptual barriers constrain the effectiveness of the search and threaten to introduce false positive predictors of false positive studies.

Formal specification of prior hypotheses, while empirically predictive of more valid positive findings, is an artefact, not a cause of such accuracy. In order for the hypothesis to be defined in advance and narrowly focused, for few statistical tests to be conducted, and for the study not to be categorized as a ‘fishing expedition’, the prior evidence in support of the hypothesis must be sufficiently strong. The biological context, experimental support, or prior epidemiological studies presumably lay the foundation that enables the researcher to specify a hypothesis for evaluation. The act of articulating the hypothesis obviously does not magically confer improved quality to the study. The prior evidence in support of the hypothesis simultaneously

enables the investigator to focus the study and makes observed positive findings more likely to be correct.

Every single hypothesis which a study addresses, even if there are hundreds of such questions, has some level of credibility prior to the study (whether the investigator knows it or not) and a new level of support after the study is completed. Studies do not generate new hypotheses; they only provide evidence that helps to evaluate the credibility of hypotheses. If the hypothesis did not exist before the study was conducted, there would have been no reason to calculate the measure of effect. To counter the notion that epidemiological studies ‘generate’ hypotheses, as opposed to providing evidence to evaluate them, Philip Cole laid the question of prior hypotheses to rest for all epidemiologists through his hypothesis generating machine. By proposing every imaginable combination of exposure and disease, all hypotheses from that point forward can be considered to have been formally specified a priori, whether or not the investigators knew it or made any use of the information. To my knowledge, with all studies from 1993 forward now based on a priori hypotheses, there was no discernible improvement in the quality of epidemiological evidence, suggesting that the prior specification of hypotheses did not in fact generate more accurate results.

It is the pre-existing support from epidemiology and other disciplines that enhances the probability that a new positive result will be a true positive, in that it would be consistent with prior evidence. When studies of identical quality, undertaken with little or no prior supporting evidence (referred to as the ‘pursuit of improbable hypotheses’ by Hatch) generate positive findings, this new positive evidence is by definition incompatible with the lack of prior support or may contradict previous findings. In those situations, the cumulative support for the hypothesis continues to be quite modest, so that the positive study is operationally defined as a ‘false positive’. Bayesian inference formalizes the prior belief in a hypothesis and the influence of a given study in shifting that belief in one direction or the other to a degree that depends on the quality and precision of the study. When the prior evidence for a hypothesis is strong, a positive study is called a ‘true positive’ study and when the prior evidence is weak or in the opposite direction, the positive study is called a ‘false positive’ study. The mistake is to confuse an increment in support from a positive study with cumulatively strong support for the hypothesis. In so-called ‘fishing expeditions’, many hypotheses, each with very limited prior support, are being evaluated.

The other correlates of prior specification of hypotheses suggest that studies with a focused hypothesis are undertaken with more methodological rigor. Thus, the use of cancer registry data or having access to the tremendous data resources available in Scandinavia, which reduce the barrier to conducting research, are predictive of false positive studies. Studies that are more expensive and difficult to initiate, reflected by cohort designs and adjustment for smoking and other confounders, are less vulnerable to false positive findings. Obtaining research funding generally requires a focus on specific hypotheses, with the focus allowing more rigorous measurement of key variables, for example. The declaration of the hypothesis in advance does not convey benefit except insofar as it leads to a methodologically stronger study. In fishing expeditions, the quality of the evidence bearing on the many hypotheses is often weak, given the failure to focus energy and resources on any one of the many questions being addressed.

Thus, I hypothesize that the benefits of specifying prior hypotheses are spurious, once the level of prior supporting evidence and the quality of the study methods are considered. Even though there may well be predictive value, in that positive results from a focused study are more likely to qualify as ‘true positives’ than similar results from a study that addresses many questions, this is a result of confounding by prior evidence and study quality. To ignore these correlates of hypothesis specification would lead to the suggestion that we simply articulate our goals or set few as opposed to many goals to avoid making errors. A colleague once suggested that the validity of positive findings could be enhanced if a small number of hypotheses were written in advance, placed in a sealed envelope, and opened upon completion of data collection and analysis. Swaen et al.’s proposed automatic downgrading of findings from fishing expeditions likewise suggests we need only state which fish we are looking for in order to enhance the value of the fish we catch.

Both the prior evidence and quality of the study deserve intense scrutiny, and the more fully each is understood, the more effectively the cumulative evidence can be characterized. When a hypothesis with limited prior support is being addressed, even a strikingly positive study remains likely to be a false positive in that the study is positive but the cumulative evidence for the hypothesis is negative. The study’s internal strengths will determine how much it shifts the evidence for or against the hypothesis in the direction of its findings. It is the cumulative evidence, not the shift in evidence from the new study, which should serve as the basis for individual decision making and setting policy. Most positive findings from epidemiology really do call only for ‘more research’.

References