LETTER

On science, statistics and mental states

Eyal Shahar, MD, MPH

Professor, Division of Epidemiology and Biostatistics, Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, AZ, USA

Correspondence address
Dr. Eyal Shahar, Division of Epidemiology and Biostatistics, Mel and Enid Zuckerman College of Public Health, The University of Arizona, 1295 N. Martin Avenue, Tucson, AZ 85724,USA. E-mail: shahar@email.arizona.edu

Accepted for publication: 17 October 2011.

To the Editor

Having read Penston's book [1], I am familiar with his arguments and viewpoints, some of which were recently published in this journal [2] and elsewhere [3]. Rather than duplicating my commentary on his book [4], I will try to add new insight into the debate [5-7].

I have come to the conclusion that much ink and many bytes could have been saved if more people understood what bias and variance mean in research. I mean "understood *exactly* what they mean", not "sort of". Unfortunately, that would also require a solid understanding of several other terms, not all of which are statistical: determinism, indeterminism, causal parameter, effect modification, estimator, estimate and distribution [8]. But that's about it - I promise. Believe it or not, almost all of science is founded on these terms. If all of us understood them, there would be no disagreement on the meaning of "knowledge", except perhaps on the axioms of science [9]. For instance, the fuzzy ideas of internal validity and external validity [8] would not be needed. The former will be replaced with confounding bias or some other bias and the latter will be replaced with effect-modification bias [10]. Moreover, the connection between an estimated effect (say, an estimated probability ratio) and a single treatment decision will be clear and conjectural [11], rather than obscure or nonexistent [2].

For the time being, I will offer a few comments on Penston's articles [2,3] and two comments on Berger's letter [5]. Miller's letter [6] sides with Penston's letter [7] on the (lack of) merit of large randomized trials.

If I understand Penston correctly, he is looking for a miraculous method that is sufficient to identify a causal relationship. He would like to be sure that a measure of association is a measure of effect. Unfortunately, he is looking for certain knowledge that never existed and will never exist. For example, we can never exclude another unknown confounder, not even in a mega randomized trial [10, 12]. So the futile quest for certainty should be replaced with two questions: What are the sources of our *conjectural* knowledge? What may we do to acquire that knowledge? The answer to the first question was given by Karl Popper many years ago: "There are all kinds of sources of our knowledge; but *none has authority*." I am intrigued to read Penston's answer to the second question. No method he might propose (including lab experiments on "identical" mice) provides certain knowledge. And the opposite of certain knowledge is conjectural knowledge - not reliable knowledge, probable knowledge, justified knowledge, or any other psychology-based adjective. The adjective "conjectural" describes logical reality (it is impossible to know that we have hit on the truth); the competitors are descriptors of wishful thinking or a mental state. As far as science is concerned, neither is interesting. You may be very confident that you should treat ischemic stroke with tPA, but don't dare to inject that drug into my vein ("within the window of opportunity"), should I suffer an ischemic stroke. I have read the same literature that you have read and don't share your confidence. And please, don't waste your time on trying to impose your mental state on mine. Leave persuasive rhetoric to politicians and salespersons. That's their craft.

Penston thinks that large studies are an indication of weakness. If their goal is to issue a verdict on the null hypothesis, the extra effort is indeed useless. I have no interest in rejection of the null and I think that most scientists would agree with me once they understood that rejection of the null neither validates the point estimate nor adds meaningful knowledge [13]. Nonetheless, I prefer large studies over small ones because sample size is a relevant property of a study. I prefer a study in which the variance of the distribution to which the estimate belongs is smaller - just as I prefer a study in which some confounding path was blocked. Make no mistake: that's a preference for a method, not "more confidence" in the estimate.

Time and again, people seem unable to grasp the essence and limits of science: we prefer one method to another not because the former will deliver the truth and the latter will not. We simply have nothing better to pursue other than a better method. But a better method does not imply that the *result* is better. I repeat: A better method does not imply that the *result* is better. For example, the realized 99% confidence interval [1.9, 2.1] does *not* mean, of course, that the parameter, θ, is located in the interval. (And it is senseless to say - in the frequentist world - that we are 99% confident that $1.9 \leq \theta \leq 2.1$.). Why don't they teach this painful truth in STAT101, instead of another statistical test?

Penston places medical research and physics at two ends of the spectrum. I have bad news to share with him and with others who think of physics as "hard science" and of epidemiology as "soft science". Physics experiments might be able to reduce information bias and confounding bias better than epidemiological studies, but physics is loaded with never-ending math, and it's difficult to tell where the math ends and the science begins. I suggest that physicists give some thought to thought bias [14] and ask themselves how much of that bias is entrenched in their science. It is not at all clear to me which is more scientific: a full page of definitions and derivations in thermodynamics, or a single odds ratio from a case-control study?

Berger took upon himself the difficult task of defending statistics. He tries to separate statistics from the statisticians, arguing that we should not dispose of statistics just because there are bad statisticians. I agree. We should not dispose of science just because there is no shortage of poorly trained scientists [15]. But there are two problems in his arguments. First, instead of defending statistics, Berger defends a study design called a randomized trial. A randomized trial belongs to statistics as much as a case-control study does. With all due respect, let's restrict the term "statistics" to math [4]. More important, which statistics does he promote as an asset to medical research? Fisherian null hypothesis testing? Neymanian null hypothesis testing? Bayesian null hypothesis testing? Credible intervals? Confidence intervals? Likelihood intervals? Marginal structural models? Are all of them assets to be picked from the shelf? Have statisticians already stopped fighting among themselves on the "right" statistics - say, Bayesian versus frequentist analysis of a randomized trial? Not to my knowledge. Too many statisticians make a living from their naïve customers who crave for a "statistically significant" result (and don't have a clue what it means [13]). Others are 100% confident that we should all be Bayesians (although they haven't agreed yet on the "right" priors).

I am a strong believer in the merit of destructive criticism [8, 9, 15, 16], but I will end with a constructive suggestion. Let's develop an honest view of science and statistics and let's start teaching that view as soon as we start teaching "science": maybe at age 10. Then, I will not have to reject the naïve perception of science and statistics ($P<0.05$) that prevails in the minds of my graduate students.

References

[1] Penston, J. (2010). stats.con: How we've been fooled by statistics-based research in medicine. London: The London Press, UK.
[2] Penston, J. (2011). The irrelevance of statistics-based research to individual patients. *International Journal of Person Centered Medicine* 1 (2), 240-249.
[3] Penston, J. (2011). Statistics-based research – a pig in a poke? *Journal of Evaluation in Clinical Practice* 17, 862–867.
[4] Shahar, E. (2011). Research and medicine: human conjectures at every turn. *International Journal of Person Centered Medicine* 1 (2), 250-253.
[5] Berger, V.W. (2011). Trials: The worst possible design (except for all the rest). *International Journal of Person Centered Medicine* 1 (3), xxx-xxx.
[6] Miller, C.G. (2011). Evidence-based medicine or ignorance-based evidence? *International Journal of Person Centered Medicine* 1 (3), xxx-xxx.
[7] Penston, J. (2011). There is no baby in the bath water. *International Journal of Person Centered Medicine* 1 (3), xxx.
[8] Shahar, E. (2008). Does anyone know the road from a randomized trial to personalized medicine? A review of 'Treating Individuals. From Randomized Trials to Personalised Medicine'. *Journal of Evaluation in Clinical Practice* 14 (5), 726-731.
[9] Shahar, E. & Shahar, D.J. (2011). Marginal structural models: much ado about (almost) nothing. *Journal of Evaluation in Clinical Practice*, published online 24 Aug, 2011 [DOI:10.1111/j.1365-2753.2011.01757.x]
[10] Shahar, E. & Shahar, D.J. (2012). Causal diagrams and three pairs of biases. In: Epidemiology (Lunet, N., Editor, ISBN 979-953-307-498-5), INTECH: Open Access Publisher, http://www.intechweb.org/ (forthcoming).
[11] Shahar, E. (2007). Estimating causal parameters without target populations. *Journal of Evaluation in Clinical Practice* 13, 814-816.
[12] Shahar, E. & Shahar, D.J. (2009). On the causal structure of information bias and confounding bias in randomized trials. *Journal of Evaluation in Clinical Practice* 15, 1214-1216.
[13] Shahar, E. (2007). Commentary: Interpreting the interpretation. *Journal of Evaluation in Clinical Practice* 13, 693-694.
[14] Shahar, E. & Shahar, D.J. (2010). Causal diagrams, information bias, and thought bias. *Pragmatic and Observational Research* 1, 33–47.
[15] Shahar, E. (2009). Commentary: Making sense of health statistics? *Journal of Evaluation in Clinical Practice* 15, 1217–1219.

[16] Shahar, E. (1998). Evidence-based medicine: a new paradigm or the Emperor's new clothes? *Journal of Evaluation in Clinical Practice* 4, 277-282.