Vandenbroucke\(^1\) comments ‘Observational research that is started with equally high priors can be expected to perform as well as randomized trials.’ Empirical evidence is on the other side. It is not just hormone replacement therapy that has come a cropper. What about vitamins E and C, and selenium and multivitamins, to name but a few?\(^2-4\) Mazik\(^5\) points to what appears to be a general problem: claims coming from observational studies generally fail to replicate. See also Ioannidis\(^6\) and Boffetta \(\textit{et al.}\)\(^7\) One small step forward would be to acknowledge that multiple testing is likely part of the problem. How can you ask tens to hundreds of questions and not expect false positives? Those that argue for no correction for multiple testing\(^8,9\) leave me not persuaded.

One purpose of a scientific paradigm is to lead to good practice and empirically verified results. Why continue a data analysis strategy that is wrong 80–90% of the time? Multiple testing should be fixed before someone says, why fund any of this research?

References

\(^1\) Vandenbroucke JP. Commentary: Mazik’s essay, seen from another angle. \textit{Int J Epidemiol} 2009;\textbf{38}: 410–12.


\(^8\) Rothman KJ. No adjustments are needed for multiple comparisons. \textit{Epidemiology} 1990;\textbf{1}: 43–46.


\texttt{doi:10.1093/ije/dyp188}

Advance Access publication 17 April 2009

Author’s Response

From JAN P VANDENBROUCKE

Leiden University Medical Centre, The Netherlands. E-mail: j.p.vandenbroucke@lumc.nl

Dr Young has one cogent argument: he is ‘not persuaded’ by those that argue for no correction for multiple testing. Fine, but what are his reasons? What is, for example, the difference between genome-wide testing of the association of a few hundred thousand SNPs with disease, and a PhD student who uses data collected for another purpose to see whether new findings by other researchers hold? The latter student is also part of the game of asking ‘hundreds of questions’. The main examples that Young cites, Vitamin C, E, selenium etc. have almost certainly nothing to do with the multiple testing problem. The statement that observational studies ‘generally fail to replicate’ is unsubstantiated. Moreover, failure of replication is expected when explanatory science develops, and should be more frequent for etiologic research than for randomized trials, as I have argued in the paper that left Dr Young not persuaded.\(^1\) Finally, generalizations about observational research are unwarranted: all genetics is observational research, so are all infectious disease outbreak investigations, all study of prognosis is observational, as are most studies of diagnosis. It is more meaningful to think separately about different fields of observational research—e.g. for adverse effects research, there is no systematic difference between randomized and observational evidence.\(^2\) That is at least one exception to Young’s allegation that observational research uses analytic strategies that are ‘wrong 80–90 percent of the time’.