

- ²⁵ Collins R. What makes the UK Biobank special. *Lancet* 2012;**379**:1173.
- ²⁶ Swanson JM. UK Biobank and selection bias. *Lancet* 2012;**380**:110.
- ²⁷ Davey Smith G, Lawlor DA, Harbord R *et al*. Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. *PLoS Med* 2007;**4**:e352.
- ²⁸ Mayer-Schonberger V, Cukier K. *Big Data*. London: John Murray, 2013.
- ²⁹ Mills H. Analytics: turning data into dollars. *Forward Focus*. 1 December 2011; http://www.deloitte.com/assets/Docom-UnitedStates/Local%20Assets/Documents/FSI/US_FSI_Forward%20Focus_Analytics_Turning%20data%20into%20dollars_120711.pdf (15 May 2013, date last accessed).
- ³⁰ Davey Smith G. Epidemiology, epigenetics and the 'Gloomy Prospect': embracing randomness in population health research and practice. *Int J Epidemiol* 2011;**40**:537–62.
- ³¹ Davey Smith G, Ebrahim S. Epidemiology—is it time to call it a day. *Int J Epidemiol* 2001;**30**:1–11.
- ³² Stallones RA. To advance epidemiology. *Ann Rev Public Health* 1980;**1**:69–82.
- ³³ Bell JT, Saffery R. The value of twins in epigenetic epidemiology. *Int J Epidemiol* 2012;**41**:140–50.

Rebuttal: When it comes to scientific inference, sometimes a cigar is just a cigar

Kenneth J Rothman,^{1,2*} John EJ Gallacher³ and Elizabeth E Hatch²

¹RTI Health Solutions, RTI International, Research Triangle Park, NC, USA, ²Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA, and ³Institute of Primary Care and Public Health, Cardiff University, Cardiff, UK

*Corresponding author. RTI Health Solutions, Research Triangle Institute, 200 Park Offices Drive, Research Triangle Park, NC, USA. E-mail: krothman@rti.org

Accepted 23 May 2013

We are grateful to the editors for suggesting that our submission become a debate piece, as we value critical discussion. We are gratified that the three invited counterpoints not only agree with our position but add useful insights. Elwood summed up our view when (referring to the White Paper on the U.S. National Children's Study¹) he commented that 'the concept of external validity given confuses statistical inference with scientific inference'.² Richiardi *et al.* echoed our point that representativeness is not desirable even if the goal is to study effect-measure modification: 'Similarly using non-representative samples may enhance our ability to assess heterogeneity with regards to potential effect modifiers, e.g. by ensuring that there are adequate numbers in each of the ethnic groups to be considered if we suspect or are interested in potential modification by ethnicity'.³ And we especially liked Nohr and Olsen's quotable remark that 'Representativeness is time and place specific and will therefore always be a historical concept...'.⁴

Richiardi *et al.* suggested that 'Perhaps Rothman and colleagues go too far in arguing that representativeness

should be avoided as a matter of principle, and we consider that there are some situations where representativeness is the most sensible approach. For example, it would be rare for researchers to only study one age-group, and to then attempt to extrapolate their findings to other age-groups, if sufficient numbers and funding were available to also sample adequate numbers from these other age-groups'. But we in fact acknowledged that there is a role for representativeness in certain circumstances, as when 'public-health professionals may rely on representative samples to describe the health status of specific populations'.⁵ Nevertheless, when studying effects across a range of a variable such as age, representativeness is not the most effective way to do so, as Richiardi *et al.* themselves stated.³ We also note that representativeness can mitigate the problem that historically some groups, such as women, children and minorities, have been underrepresented or omitted from research studies. Sampling representativeness, however, is not necessary to fix that problem. Deliberate oversampling of the understudied groups would do so, and be scientifically more efficient.

Thus, despite some quibbles, all three of the counterpoints are well aligned with our contention that representativeness, although it may have a place in health surveys, is not a proper goal for scientific studies. The only real disagreement comes from Ebrahim and Davey Smith,⁶ who expressed concern regarding this emerging consensus. We believe, however, that their defence of the classical view of representativeness does not clarify our understanding of the topic.

Ebrahim and Davey Smith portray us as 'elevating' causal hypothesis testing to the level of science and 'denigrating' descriptive epidemiology as 'not science'.⁶ Our point was simply that not everything that an epidemiologist does is science, in the sense that not every activity of epidemiologists adds to the inventory of statements describing how nature works. We did acknowledge that some studies should use representative sampling. We also said that these studies are 'informative, but they are not science in the same way that causal studies about how nature operates are science'.⁵ Yes, we were making a distinction between theoretical and applied work, but references to 'elevation' of causal studies and 'denigration' of population health surveys come from Ebrahim and Davey Smith, not from us. Distinguishing one set of goals from another does not necessarily imply a hierarchy. When Karl Popper used falsifiability as a demarcation criterion to distinguish scientific statements from metaphysical statements, the distinction was not an attempt to denigrate metaphysics. We consider the placing of the Mars rover on Mars to be a stupendous engineering achievement, but to describe it as engineering and not science is not a slur, and does not detract from the achievement. Sometimes a cigar is just a cigar.

Ebrahim and Davey Smith suggested that non-representative study groups may produce biased associations. Their argument is based on a shaky premise, namely that bias is judged by the difference between the expected study results and the population average, as if the population average represents a unique truth. If the association being measured varies by subgroup, the overall association in a study that is weighted differently from the source population is simply a summary that is weighted to a different standard. If replicating the association in the source population is desired, the investigators can standardize to that distribution. They can do so without sampling subjects according to that weighting, and may have good reason to weight differently from the source population, for example, to maintain study efficiency. The point is that non-representativeness does not produce biased results simply because the study population differs by age or some other factor from a designated target population, unless we circularly define the results to be biased because they are non-representative. It is fine to pursue representativeness when it is needed, as in a health survey, but lack

of representativeness does not automatically amount to bias.

We of course agree that selection factors can lead to selection bias (which we would define as systematic errors stemming from procedures or factors involved in subject participation). For the bias affecting the relation to alcohol consumption and stroke in the American Cancer Society study⁷ to materialize, Ebrahim and Davey Smith theorized that the study participants may have comprised heavy drinking epidemiologists with a low risk of stroke, because they smoke less, exercise heavily and exhibit health consciousness, but the participants who were lighter drinkers smoked more, exercised less and showed less health awareness, and consequently had a comparatively high risk of stroke. It seems unlikely that these assumptions could account for the magnitude of bias that Ebrahim and Davey Smith attribute to them, but even if this dubious situation did occur, representative sampling would tend to obscure rather than solve the problem, as the study would still comprise volunteers and health awareness would still be a possible selection factor distinguishing study participants from those who choose not to participate. The bias could be controlled, with or without representative sampling, by measuring and controlling for health awareness, using information about health-seeking behaviour such as medical screening visits, influenza vaccinations and other indicators of the selection factor underlying their concern. Behaviours such as smoking and exercise, which Ebrahim and Davey Smith name as sources of the supposed bias, were actually measured and controlled in the American Cancer Society study.⁷

We think it unwise to assume, as Ebrahim and Davey Smith have done in their discussion of studies of vitamin C and antioxidants, that any finding from a randomized trial must be valid and would overrule contradictory findings from non-experimental studies. For example, no one seems to think that a randomized trial that showed that smoking cessation was associated with an increase in lung cancer⁸ should trump evidence on smoking and lung cancer from other sources, which include the cohort study of 35 000 highly selected, male, UK doctors who were followed for decades.⁹ Trials also harbour sources of uncertainty and error, and to argue in favour of representativeness because some cohort studies based on volunteers obtain results that differ from those of selected trials (also based on volunteers) is simplistic. Reconciling conflicting results is challenging, but one should be wary of simply ascribing the differences to study design, as Hernán illustrated in his reconciliation of the conflicting results from trials and cohort studies reporting the relation between hormone replacement therapy and cardiovascular mortality.¹⁰

Ebrahim and Davey Smith argue that lack of representativeness leads to greater confounding in

epidemiological studies ('many variables of interest will be associated with participation and essentially volunteer samples may suffer from greater degrees of confounding than less selected samples'⁶). One of our points was that traditional design of experiments involves holding potential confounding factors constant, to prevent confounding. In contrast, Ebrahim and Davey Smith appear to believe that confounding is reduced by replicating the population associations within the study population, rather than holding variables constant. But that would merely reproduce in the study population whatever confounding exists in the parent population. Consider age, for example: perfect representative sampling will not reduce age confounding in a study if age is confounding in the source population. In contrast, restriction by age would do so. True, restriction would prevent studying the extent to which age is an effect-measure modifier, but as stated earlier, if that is a study goal, representative sampling is an inefficient way to achieve it.

These arguments are not merely academic. Epidemiological studies focused on causal explanations are often large and expensive and can do without the burden of attempting to achieve representativeness. As the U.S. National Children's Study illustrates, the futile pursuit of representativeness where it is not needed can be extraordinarily costly. We acknowledge that descriptive studies that are intended to be applicable to specific populations at a specific time may need to pursue representativeness. The latter are specialized studies, whose design should not carry over into aetiological research. A clear separation of these two aims will reduce unnecessary costs and lead to better studies. Fortunately, all the writers contributing to this discussion seem to agree on some basics. Even Ebrahim and Davey Smith, despite considerable areas of disagreement, offer a concluding sentence that could easily have characterized our position: 'We feel that representativeness should neither be avoided nor uncritically universally adopted, but its value evaluated in each particular setting'.⁶

Acknowledgement

We are grateful to Lauren Wise for helpful suggestions.

Funding

K.J.R. and E.E.H. were supported by grant # R01 HD-060680 from the National Institute of Child Health and Human Development. J.E.J.G. was supported by funding from the UK Biobank.

Conflict of interest: None declared.

References

- ¹ Rust S. *White Paper on Evaluation of Sampling Design Options for the National Children's Study*. Appendix A: Advantages and limitations of probability-based sampling for the NCS. Available at <http://www.nationalchildrensstudy.gov/research/whitepapers/Pages/default.aspx>; <http://www.nationalchildrensstudy.gov/research/whitepapers/Pages/Appendix-A-Advantages-and-Limitations-of-Probability-Based-Sampling-for-the-National-Children-s-Study.pdf> (16 July 2013, date last accessed).
- ² Elwood JM. On representativeness. *Int J Epidemiol* 2013;**42**:1014–15.
- ³ Richiardi L, Pizzi C, Pearce N. Representativeness is usually not necessary and often should be avoided. *Int J Epidemiol* 2013;**42**:1018–22.
- ⁴ Nohr EA, Olsen J. Epidemiologists have debated representativeness for more than 40 years—has the time come to move on? *Int J Epidemiol* 2013;**42**:1016–17.
- ⁵ Rothman KJ, Gallacher J, Hatch EE. Why representativeness should be avoided. *Int J Epidemiol* 2013;**42**:1012–14.
- ⁶ Ebrahim S, Davey Smith G. Should we always deliberately be non-representative? *Int J Epidemiol* 2013;**42**:1022–26.
- ⁷ Thun MJ, Peto R, Lopez AD *et al*. Alcohol consumption and mortality among middle-aged and elderly U.S. adults. *N Engl J Med* 1997;**337**:1705–14.
- ⁸ Shaten BJ, Kuller LH, Kjelsberg MO *et al*. Lung cancer mortality after 16 years in MRFIT participants in intervention and usual-care groups. *Ann Epidemiol* 1997;**7**:125–36.
- ⁹ Doll R, Peto R, Boreham J, Sutherland I. Mortality in relation to smoking: 50 years' observations on male doctors. *BMJ* 2004;**328**:1519–28.
- ¹⁰ Hernán MA, Alonso A, Logan R *et al*. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology* 2008;**19**:766–79.