However, it should be emphasized that, regarding selection, the issue can be solved by taking into account the temporal relationships between the variables under study and, thus, by enrolling the participants before the intermediate variable or its early signs could become manifest. In a birth cohort study involving enrolment during the first trimester of pregnancy, for example, selection cannot be directly affected by intermediate variables acting later in pregnancy or at birth.

## Conclusions

In conclusion, we agree with Rothman and colleagues that scientific inference does not require representativeness, and often explicitly requires that study samples should not be representative. Overall, representativeness can be harmful or beneficial depending on the study question and context. There is no reason to embrace representativeness per se, as often restriction can enhance the practicality of a study and/or the validity of the scientific inferences. We acknowledge that further work is needed to fully understand some specific situations, in particular when an intermediate variable directly affects baseline selection. However, leaving aside this specific issue, we consider that the view that studies based on representative samples are clearly better than those based on restricted samples is untenable. Rather, although it is perhaps too strong to argue that representativeness should always be avoided, it is usually not necessary, and often should be avoided.

**Conflict of interest:** None declared.

## References


1. Rothman KJ, Gallacher J, Hatch EE. Why representativeness should be avoided. *Int J Epidemiol* 2013;**42:**1012–14.
2. Doll R, Hill AB. The mortality of doctors in relation to their smoking habits: a preliminary report. *Br Med J* 1954; **1:**1451–55.
3. Kolonel LN, Henderson BE, Hankin JH *et al*. A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. *Am J Epidemiol* 2000;**151:**346–57.
4. Asher MI, Keil U, Anderson HR *et al*. International Study of Asthma and Allergies in Childhood (ISAAC): rationale and methods. *Eur Respir J* 1995;**8:**483–91.
5. Richiardi L, Baussano I, Vizzini L, Douwes J, Pearce N, Merletti F. Feasibility of recruiting a birth cohort through the Internet: the experience of the NINFEA cohort. *Eur J Epidemiol* 2007;**22:**831–37.
6. Doll R, Hill AB. Mortality in relation to smoking: ten years' observations of British doctors. *Br Med J* 1964;**1:** 1399–410.
7. Pizzi C, De Stavola B, Merletti F *et al*. Sample selection and validity of exposure-disease association estimates in cohort studies. *J Epidemiol Community Health* 2012;**65:** 407–11.
8. Pizzi C, De Stavola BL, Pearce N *et al*. Selection bias and patterns of confounding in cohort studies: the case of the NINFEA web-based birth cohort. *J Epidemiol Community Health* 2012;**66:**976–81.
9. Howards PP, Schisterman EF, Poole C, Kaufman JS, Weinberg CR. ''Toward a clearer definition of confounding'' revisited with directed acyclic graphs. *Am J Epidemiol* 2012;**176:**506–11.


Published by Oxford University Press on behalf of the International Epidemiological Association

*International Journal of Epidemiology* 2013;**42**:1022–1026
doi:10.1093/ije/dyt105

# Commentary: Should we always deliberately be non-representative?

**Shah Ebrahim[1]* and George Davey Smith[2]**

[1]Department of Non-communicable Disease Epidemiology, London School of Hygiene & Tropical Medicine, London,UK and [2]MRC Centre for Causal Analyses in Translational Epidemiology, University of Bristol, Bristol, UK

*Corresponding author. Department of Non-communicable Disease Epidemiology, London School of Hygiene & Tropical Medicine, Keppel Street, London WC1E 7HT. E-mail: Shah.ebrahim@lshtm.ac.uk

**Accepted** 1 May 2013

Rothman and colleagues were invited to submit their piece to our recently established 'Education Corner', but on reading it we felt it merited discussion and debate.[1] Those invited to comment considered that

Rothman *et al*. had overstated their position but were basically correct.[2–4] We are concerned that this notion will become accepted wisdom in epidemiology without its implications having been thought through, and feel that representativeness should neither be avoided nor uncritically embraced, but adopted (or not) according to the particular questions that are being addressed.

## Some uses of epidemiology require representative samples

The purpose of epidemiology is not simply to assess causal hypotheses.[5] Rothman *et al.* elevate causal hypothesis testing to 'a science' and denigrate descriptive epidemiology as an applied practice and therefore 'not science'. It is salutary to reflect that the main reason millions of dollars are poured into modelling the global burden of disease[6] (GBD) is because epidemiologists neglected their responsibility to collect data on prevalence and incidence of disease from defined populations. Whatever one's views are of the GBD initiative, it is certainly the case that evidence-based healthcare planning and policy require population-representative data.

## Non-representative study groups may produce biased associations

In causal studies, the fundamental concern is that lack of representativeness will introduce bias. Richiardi *et al.* in their commentary suggest that non-representative populations produce only weak bias in exposure-disease associations.[2] Their conclusion is based on two of their own studies, one empirical and the other theoretical. Their empirical assessment compares associations of educational attainment and parity with two outcomes: low birthweight and caesarean section. They claim that non-representative internet sampling 'does not necessarily introduce selection bias'. However, reviewing the data (Pizzi *et al.*'s Table 4[7]) indicates that the non-representative sampling resulted in odds ratios that were lower for most comparisons, and the confidence intervals (CIs) of the relative difference in odds ratios between non-representative and representative sampling were consistent with effect sizes that would excite interest in many epidemiologists and result in a media frenzy. For example: 'Any red meat you eat contributes to the risk (of death)', claimed the lead author of a paper from the Health Professions and Nurses Health studies based on a relative risk estimate of 1.20, well within the limits of what was observed in these empirical estimates of the possible bias.[8] Similar excitement was generated by a study that demonstrated a relative risk of 1.06 (95% CI 1.02–1.10) for cancer mortality in relation to a very large difference in intake of processed meat per day.[9] The important point is that the distortions may be in any direction (which is unpredictable), and evidence from one empirical example does not necessarily apply to other causal questions.

Richiardi *et al.*'s theoretical example[10] uses Monte Carlo simulations to study the effects of selection into a study, claiming that the bias is small, particularly for relative risks between 0.5 and 2.0. However, they only examined scenarios where a single unmeasured determinant of the outcome also influenced the selection process, as they believe that 'it is unlikely that multiple and independent important disease risk factors would affect the sample selection'.[10] This is a surprising belief for these authors to hold, as their own empirical findings (admittedly published after this paper) show very clearly that multiple risk factors are indeed associated with participation (Table 1 in[7]).

Why might the issue of multiple factors being associated with participation be important? Take the example of vitamin C levels and coronary heart disease (CHD) events: a large body of observational data from studies conducted at different times and places has produced a precise and repeatable estimate of an apparent benefit from higher vitamin C levels.[11] But exploration of the complex confounding of the association demonstrates how multiple factors, operating across the life course, can lead to confounding strong enough to negate the apparent benefit.[12–14] A large randomized controlled trial of vitamin C supplementation, in which such confounding should not arise, showed no strong evidence of any reduction in CHD events.[15] A similar scenario, with conflicting results from observational and experiments, has been seen with respect to other antioxidant vitamins.[16] The simple fact is that multiple factors do come together to generate sometimes sizeable non-causal associations, even after attempted statistical adjustment for confounding factors.

The large American Cancer Society volunteer cohort is exactly the sort of non-representative study group that Rothman and others would consider fit for purpose—it is easy to follow up, participants are motivated to stay in the study and events are likely to be easy to count. In this volunteer cohort, high alcohol consumption was associated with a reduced risk of stroke,[17] a surprising finding since the outcome included haemorrhagic stroke (for which alcohol might be expected to increase risk) and alcohol increases blood pressure which is a major causal factor for stroke.[18,19] What type of heavy drinker would volunteer for a study about the health effects of their lifestyle? They are unlikely to be representative of all heavy drinkers in the population (e.g. they may be non-smoking, vigorous exercising, moderately wealthy epidemiologists) and the factors that make them non-representative will tend to render them at lower risk of stroke. By contrast, volunteers drinking moderately or less and non-drinkers may be more representative of moderate, low and non-drinkers in

the general population. This non-representative cohort generates a potentially spurious result because many factors that are associated with the outcome of interest are also likely to be linked to self-selection into a study.

## Scientific generalization: animals and randomized controlled trials

In support of the argument for non-representative study groups, Rothman *et al.* state that scientific generalization is incongruous with representative sampling, only modestly reworking Rothman's earlier views on the topic of representativeness.[20] Using as examples animal experiments (where no attempt is made to sample from a population of animals) and randomized trials (where internal validity may be achieved by limiting recruitment to a narrowly defined group), an argument is developed that representativeness is counter-productive. The validity of animal experiments of pharmacotherapies has been widely questioned, as it has become clearer that, despite careful control for confounding factors, many of them get the wrong answers or are conducted with no intention of application in humans.[21] Randomized trials of interventions that are primarily for use in older adults with multiple morbidities have been criticized for not recruiting participants more representative of those who will be treated in the real world.[22,23] Trialists simply do not know (and certainly cannot control for them or generalize from restricted study groups) the complex confounding between age-related processes, co-existing disease and therapies, and the effects of a new intervention.

## The road to non-representative studies

Epidemiologists are not often capable of producing 'general statements on nature' and unfortunately more often report on associations in ways that imply that causal inference is being drawn.[24] Rothman *et al*. consider that the best direction for epidemiology is to set up more studies that 'control skillfully for confounding variables and thereby advance our understanding of causal mechanisms'. The UK Biobank study of 500 000 people is an outstanding example of a study (motivated initially by the desire to conduct large-scale genetic investigations) which implemented a demanding protocol in terms of measurements on participants where non-representativeness was inevitable. It is claimed for UK Biobank that 'generalisable associations of exposure with disease can be obtained without including representative samples of particular populations'.[25] The overall response rate of 5.5%[26] is not prominently displayed on the UK Biobank website presumably because it is deemed irrelevant. From a purely gene-outcome association point of view, the study will, for the most part, be capable of yielding unbiased estimates of association, as genetic variants are unlikely to be associated with self-selection into the study and are not generally associated with confounding factors.[27] The large sample size, relatively cheaply recruited, is a major advantage here. Non-genetic associations will have to be interpreted cautiously, as many variables of interest will be associated with participation and essentially volunteer samples may suffer from greater degrees of confounding than less selected samples. Once this data set (and all the others) is turned over to open access, it is inevitable that large numbers of environmental variable-outcome associations of small effect size—but very precisely estimated (see the confidence intervals on the relative risk of 1.06 for processed meat cited earlier)—will be published. It is this context that Rothman *et al.*'s hope for skillful control for confounding variables reflects optimism of a Panglossian scale: in many situations the degree of unavoidable measurement imprecision and inevitable unmeasured confounders renders reliable control unattainable.

## Epidemiology in the big data world

Rothman's and colleagues' stance bears some affinity with the justifiable excitement about the big data era we are entering.[28] Proponents of this brave new world denigrate representative sampling in a way Rothman and colleagues would presumably applaud: 'Reaching for a random sample in the age of big data is like clutching at a horse whip in the era of the motor car'.[28] However, the promise of big data is explicitly not to identify causes; indeed, the 'Ideal of identifying causal mechanisms is a self-congratulatory illusion'.[28] We are in the realm of prediction, an example being the use of hundreds of variables from amount of TV people watch and the websites they visit to predict insurance risk. It's much cheaper than the lab tests insurance companies often use, and does just as well, an exercise in 'turning data into dollars'.[29] But epidemiologists are surely not yet ready to abandon the difficult business of characterizing causality.

Does any of this matter? Science is meant to be self-correcting; misleading findings will be exposed by further studies. Unfortunately, most of the studies that are capable of producing such findings share similar confounding structures (many components of which are not measured) and are only capable of making more precise but essentially meaningless estimates. Epidemiology that searches for causes of smaller and smaller effect sizes may become increasingly irrelevant when there is high profile contradictory evidence (including that pitting purely observational vs contradictory randomized controlled trial or genetic evidence). Growing awareness among the public, and research funders, of the impossibility of the

robust detection and unlikely impact on public health of such findings might lead to less support for their generation. Moreover, the importance of a stochastic element to disease risk that is not epidemiologically tractable at the individual level is now apparent and argues for a re-orientation of epidemiology away from attempts to improve prediction of individual risk or search for non-existent additional causes, and towards making good use of genetic variation which may tell us about population-level modifiable causes of common diseases.[30]

In our first editorial for *IJE* in 2001[31] we quoted Reuel Stallones who in 1980 had memorably detected a 'Continuing concern for methods, and especially the dissection of risk assessment, that would do credit to a Talmudic scholar and that threatens at times to bury all that is good and beautiful in epidemiology under an avalanche of mathematical trivia and neologisms'.[32] In the pre-modern epidemiology world the focus was often on triangulating evidence from across as many sources as possible. Such evidence comes from a variety of sources, of which some are deliberately non-representative (for example the considerable value of twin studies for identifying potentially causal epigenetic influences on disease,[33] or the follow-up of natural experiments) and some of which (including essentially 100% coverage population linkage studies) will be representative. Among these sources, large volunteer studies such as UK Biobank will be powerful tools, but will need to be combined with other approaches that allow strengthening of causal inference in observational data. We feel that representativeness should neither be avoided nor uncritically universally adopted, but its value evaluated in each particular setting.

**Conflict of interest:** None declared.

## References


[1] Rothman K, Gallacher J, Hatch E. Why representativeness should be avoided. *Int J Epidemiol* 2013;**42:**1012–14.

[2] Richiardi L, Pizzi C, Pearce N. Representativeness is usually not necessary and often should be avoided. *Int J Epidemiol* 2013;**42:**1018–22.

[3] Nohr EA, Olsen J. Epidemiologists have debated representativeness for more than 40 years – has the time come to move on? *Int J Epidemiol* 2013; doi:10.1093/ije/dyt102.

[4] Elwood JM. On representativeness. *Int J Epidemiol* 2013; **42:**1014–15.

[5] Davey Smith G. On the uses of 'Uses of Epidemiology'. *Int J Epidemiol* 2001;**30:**1146–55.

[6] Horton R. GBD 2010: understanding disease, injury, and risk. *Lancet* 2012;**380:**2053–54.

[7] Pizzi C, De Stavola BL, Pearce N *et al*. Theory and methods: Selection bias and patterns of confounding in cohort studies: the case of the NINFEA web-based birth cohort. *J Epidemiol Community Health* 2012;**66:** 976–81.

[8] Pan A, Sun Q, Bernstein A, Shulze M, Manson J *et al*. Red meat consumption and mortality: results from 2 prospective cohort studies. *Arch Intern Med* 2012;**172:** 555–63.

[9] Rohrmann S, Overvad K, Bueno-de-Mesquita HB *et al*. Meat consumption and mortality – results from the European Prospective Investigation into Cancer and Nutrition. *BMC Med* 2013;**11:**63.

[10] Pizzi C, De Stavola BL, Merletti F *et al*. Theory and methods: Sample selection and validity of exposure–disease association estimates in cohort studies. *J Epidemiol Community Health* 2011;**65:**407–11.

[11] Khaw K-T, Bingham S, Welch A *et al*. Relation between plasma ascorbic acid and mortality in men and women in EPIC-Norfolk prospective study: a prospective population study. *Lancet* 2001;**357:**657–63.

[12] Lawlor DA, Davey Smith G, Bruckdorfer KR, Kundu D, Ebrahim S. Observational versus randomised trial evidence. *Lancet* 2004;**364:**754–75.

[13] Khaw KT, Day N, Bingham S, Wareham N. Observational versus randomised trial evidence. *Lancet* 2004;**364:**753–54.

[14] Lawlor DA, Ebrahim S, Kundu D, Bruckdorfer KR, Whincup PH, Davey Smith G. Vitamin C is not associated with coronary heart disease risk once life course socioeconomic position is taken into account: Prospective findings from the British Women's Heart and Health Study. *Heart* 2005;**91:**1086–87.

[15] Heart Protection Study Collaborative Group. MRC/BHF Heart Protection Study of antioxidant vitamin supplementation in 20,536 high-risk individuals: a randomised placebo-controlled trial. *Lancet* 2002;**360:**23–33.

[16] Myung SK, Ju W, Cho B *et al*. Korean Meta-Analysis (KORMA) Study Group. Efficacy of vitamin and antioxidant supplements in prevention of cardiovascular disease: systematic review and meta-analysis of randomised controlled trials. *BMJ* 2013;**346:**10.

[17] 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.

[18] Lawlor DA, Nordestgaard BG, Benn M, Zuccolo L, Tybaerg-Hansen A, Davey Smith G. Exploring causal associations between alcohol and coronary heart disease risk factors: findings from a Mendelian Randomization study in the Copenhagen General Population Study. *Eur Heart J* 2013; doi:10.1093/eurheartj/eht081.

[19] Chen L, Davey Smith G, Harbord R, Lewis S. Alcohol intake and blood pressure: a systematic review implementing Mendelian Randomization approach. *PLoS Med* 2008;**5:**461–71.

[20] Rothman K. *Modern Epidemiology*. 1st edn. Boston, MA: Little Brown, 1986.

[21] Pound P, Ebrahim S, Bracken M, Sandercock P, Roberts I. Where is the evidence that animal experiments benefit humans. *BMJ* 2004;**328:**514.

[22] Harriette GC, Van Spall HGC, Toren A, Kiss A, Fowler RA. Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review. *JAMA* 2007;**297:**1233–40.

[23] Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Threats to applicability of randomised trials: exclusions and selective participation. J Health Serv Res Policy. 1999;**4:**112–1–21.

[24] Taubes G. Epidemiology faces its limits. *Science* 1995;**269:** 164–69.

[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/Dcom-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.

Published by Oxford University Press on behalf of the International Epidemiological Association
*International Journal of Epidemiology* 2013;**42**:1026–1028
doi:10.1093/ije/dyt124


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

**Kenneth J Rothman,[1,2]* John EJ Gallacher[3] and Elizabeth E Hatch[2]**

[1]RTI Health Solutions, RTI International, Research Triangle Park, NC, USA, [2]Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA, and [3]Institute 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 Study[1]) 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 is 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.