Published by Oxford University Press on behalf of the International Epidemiological Association
International Journal of Epidemiology 2013;42:1012–1014
doi:10.1093/ije/dys223

POINT COUNTERPOINT

Why representativeness should be avoided

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

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

Accepted 21 November 2012

The essence of knowledge is generalisation. That rubbing wood in a certain way can produce fire is a knowledge derived by generalisation from individual experiences; the statement means that rubbing wood in this way will always produce fire. The art of discovery is therefore the art of correct generalisation. What is irrelevant, such as the particular shape or size of the piece of wood used, is to be excluded from the generalisation; what is relevant, for example, the dryness of the wood, is to be included in it. The meaning of the term relevant can thus be defined: that is relevant which must be mentioned for the generalisation to be valid. The separation of relevant from irrelevant factors is the beginning of knowledge.
—Hans Reichenbach[1]

Why do so many believe that selecting representative study populations is a fundamental research aim for scientific studies? This view is widely held: representativeness is exalted along with motherhood, apple pie and statistical significance. For some researchers this goal can be so important that they would deem a study not worth undertaking if representativeness cannot be achieved. That was the case for two advisors to the U.S. National Children's Health Study, who resigned when the study design was changed so that representativeness was threatened.[2] We admire people who take a stand for principle over expediency, but what exactly is the principle that representativeness embodies? Here we suggest that representativeness may be essential for conducting opinion polls, or for public-health applications, but it is not a reasonable aim for a scientific study.

Within most scientific disciplines, sampling representativeness is incongruous with research goals. Immunologists doing experiments with hamsters do not dwell on getting a representative sample of hamsters. To the contrary, they select hamsters that are extremely unrepresentative because they are homogeneous, having identical genes, living in identical circumstances, and fed identical diets. The immunology of these hamsters may not be identical to that of people, but the expectation is that by controlling the characteristics and environment of the hamsters, inferences can be drawn that may generalize to people.

A simple view of generalization casts it as a process of constructing a correct statement about the way nature works. That process is uncertain, along with everything else in empirical science, but it is not an extrapolation from sample to target population. When Pasteur created the experiment that refuted the theory of spontaneous generation, he used a goose-neck flask to allow air to contact his cooling broth without letting organisms settle into the broth. His concern was to control the conditions in a precise way. Similarly, when John Snow conducted his natural experiment showing that London citizens imbibing diluted sewage were at much greater risk of cholera than those consuming water piped from upriver, he was not looking for a representative sample of London citizens. Instead he was looking for people whose characteristics and living conditions were comparable except for the source of their water consumption. Generalizing his findings was predicated on understanding the phenomenon at hand. When Doll and Hill studied the mortality of male British physicians in relation to their smoking habits,[3] their findings about smoking and health were considered broadly applicable despite the fact that their study population was unrepresentative of the general population of tobacco users with regard to sex, race, ethnicity, social class, nationality and many other variables.

Scientific generalization relates to the elaboration of the circumstances in which a finding applies. Newton's laws of mechanics explain many physical phenomena, although we now know that they are not applicable on very small scales, at high speeds or in strong gravitational fields. On a more modest level, consumption of contaminated shellfish can cause hepatitis A infection, but this relation is largely nullified by consumption of beverages containing at least 10% alcohol along with the shellfish.[4] The added knowledge about the modifying effect of alcohol is part of the generalization of the relation between

consumption of contaminated shellfish and the risk of infection with hepatitis A. It is not representativeness of the study subjects that enhances the generalization, it is knowledge of specific conditions and an understanding of mechanism that makes for a proper generalization.

It is true that statistical inference, the process of inferring from a sample to the source from which it was drawn, is greatly aided by having a representative sample. The mistake is to think that statistical inference is the same as scientific inference. Science works on the assumption that the laws of nature are constant, but if we conflate statistical inference with scientific inference we get the reverse principle, in which the results of a study are applicable only in circumstances just like those of the study itself, and applicable only to people who are just like those in the study population.

Indeed, representativeness can be counterproductive. Suppose a study is designed to examine the therapeutic efficacy of a drug. Consider three design alternatives: option A, enrol subjects between the ages of 40 and 49; option B, enrol the number of subjects needed from three age groups, 20-29, 40-49 and 60-69, to produce about equal numbers of outcomes in each of these age categories; or option C, enrol subjects with an age distribution that has been sampled to be representative of all patients with the problem the drug is intended to treat. Which design is best? The first design option will greatly limit age imbalances that could confound the results, thereby enhancing the study validity. It has the drawback, however, of informing about the effect only for subjects in a narrow range of age. Can inferences be drawn for patients of other ages? The answer depends on how much is known about the mechanism of effect. If little is known, then generalizing beyond the age range of study participants may be unwarranted. In that case, the study goal might be expanded to include how the effect varies by age. To do that, we would have to choose option B or C, and control for age imbalances through matching or in the analysis. If weighing options B and C to study how the effect varies by age, it is much better to choose option B, which allows three equally informative assessments in three distinct age ranges, rather than allowing the distribution of ages in the source population to determine the study design. The same point would apply to other potential effect-modifying variables. For example, to study how an effect varies by ethnic group or socioeconomic category, it would be preferable to choose equal numbers from the different groups, rather than select subjects in proportion to their numbers in the source population.

Clearly, representativeness does not, in and of itself, deliver valid scientific inference. If a study population is representative of some larger source population, the overall associations observed in the study population may not apply to every subgroup. The overall effect is merely an average effect that has been weighted by the distribution of people across these subgroups. Thus, if you have a sample that is representative of the sex distribution in the source population, the results do not necessarily apply either to males or to females, but only to a hypothetical person of average sex. If you want to study the extent to which an effect varies by subgroup of a third variable, you need to design the research to examine the effect by subgroups.

Representative sampling is needed to implement some study designs, such as control sampling from the source population in some case-control studies, but that sampling concern about controls is not the same as the representativeness of the study population itself. Seeking representativeness of the study population makes sense when sampling purely for descriptive purposes. Pollsters seek representative samples of their target populations to avoid polling everyone in the study population. Similarly, public-health professionals may rely on representative samples to describe the health status of specific populations. These descriptions are sampling snapshots that make no pretence of explaining how nature works. Their utility is in their description of a specific population at a point in time. Thus we draw a line between the scientific goal of understanding a phenomenon and the practical goal of applying that knowledge to specific populations. The first goal is not enhanced by representativeness, but rather depends more on tightly controlled comparisons drawn over a variety of relevant settings. It is the second goal, the application of science, that may require representative sampling. For example, from studies not involving representative samples, regular use of aspirin has been found to reduce the incidence of bowel cancer.[5] Given a polyp-related mechanism,[6] the public-health impact of aspirin chemo prevention would likely depend on the incidence or the prevalence of colorectal polyps in the target population. Measuring that impact on a target population would involve representative sampling.

Surveys of opinions, of the prevalence of disease, of habits or of environmental exposures may be informative, but they are not science in the same way that causal studies about how nature operates are science. Polls more than a few days old may become irrelevant, even if conducted with people in the same geographical area. Consequently polls are conducted in numerous places and repeated often. Prevalence surveys may also lose validity quickly over time, depending on the stability of the condition measured, and they are seldom generalizable across populations. In contrast, a scientific finding would be expected to be repeatable. One way to distinguish science from the kind of information that surveys produce is its overall applicability in space and time. Scientific statements ideally serve to describe nature in a way that is not limited to one time and one place. Although biological

principles seem to be vastly more varied than physics, and more dependent on locally varying modifying influences, the ultimate aim of biological research on humans or other species, is like that of physics, to be able to make general statements about nature. Paradoxical though it may seem, statistical representativeness leads to particular statements about the world, not general statements about nature. As initial steps, surveys may help to seed hypotheses and give a push toward scientific understanding, but the main road to general statements on nature is through studies that control skillfully for confounding variables and thereby advance our understanding of causal mechanisms. Representative sampling does not take us down that road.

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] Reichenbach H. *The Rise of Scientific Philosophy*. Bognor Regis, UK: University of California Press, 1951, p. 5.
[2] Reichenbach H. Children's Study Row. *Nature* 2012;**483:** 378.
[3] Doll R, Hill AB. The mortality of doctors in relation to their smoking habits: a preliminary report. *Br Med J* 1954; **ii:**1451–55.
[4] Desenclos JA, Klontz KC, Wilder MH, Gunn RA. The protective effect of alcohol on the occurrence of epidemic oyster-borne hepatitis A. *Epidemiology* 1992;**3:**371–74.
[5] Rothwell RM, Wilson M, Elwin CE, *et al*. Long-term effect of aspirin on colorectal cancer incidence and mortality: 20-year follow-up of five randomised trials. *Lancet* 2010; **376:**1741–50.
[6] Levine JS, Ahnen DJ. Adenomatous polyps of the colon. *N Engl J Med* 2006;**355:**2551–57.

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

International Journal of Epidemiology 2013;**42**:1014–1015
doi:10.1093/ije/dyt101

Commentary: On representativeness

J Mark Elwood

Department of Epidemiology and Biostatistics, School of Population Health, Tamaki Innovation Campus, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand. E-mail: mark.elwood@auckland.ac.nz

Accepted 15 January 2013

Most epidemiological studies—indeed, all the interesting ones—are designed to assess a potential causal relationship. There are often difficult choices in the selection of the subjects included in the study. Whether an intervention study, an observational cohort study or a case-control study, the selection of the subjects can influence both internal validity and external validity; and further, can modify the hypothesis being tested. Internal validity is the quality controlling whether a valid assessment of cause and effect can be made within the context of the study. External validity relates to the generalizability or application of this cause and effect assessment to other populations, and is clearly a secondary issue; if the study has very low internal validity, the conclusions are likely to be wrong, and so its generalizability is irrelevant.

With high internal validity, the valid assessment of the causal relationship may be widely generalizable, and does not require that the participants be representative of those to whom the new evidence will be applied. The value of good studies is in the fact that their results can be applied to very different populations, particularly in the future. Thus to choose the best treatments, physicians apply the results from internally valid studies, usually randomized trials, often done in different countries on patients diagnosed many years previously. We do not need to assume that the subjects involved in these earlier studies are representive, in a general way, of the new patient. Similarly we apply knowledge of genetics from fruit flies to humans, because the biological relationships are generalizable although the individuals studied are not. An epidemiological example is the UK Biobank cohort study: whereas