Avoidance of representativeness in presence of effect modification

From Andreas Stang1,2* and Karl-Heinz Jöckel3

1Institut für Klinische Epidemiologie, Medizinische Fakultät, Martin-Luther-Universität Halle-Wittenberg, Halle, Germany, 2School of Public Health, Department of Epidemiology, Boston University, Boston, MA, USA and 3Institut für Medizinische Informatik, Biometrie und Epidemiologie, Universitätsklinikum der Universität Duisburg-Essen, Essen, Germany

*Corresponding author. Institut für Klinische Epidemiologie, Martin-Luther-Universität Halle-Wittenberg, Magdeburger Str. 8, 06112 Halle (Saale), Germany. E-mail: andreas.stang@uk-halle.de

We read with interest the contribution by Rothman et al1,2 who believe that representativeness of a target population should be avoided. Avoidance of representativeness is an active process and obviously means that the investigator intentionally selects a subgroup of the target population that is not representative of the target population. As a result, the distributions of exposures and disease risks do not represent those of the target population any more. We believe that Rothman et al.’s suggestion is only reasonable for associations that are not modified by selection factors that modify the association between the exposure and disease of interest.

Rothman et al. state that ‘science works on the assumption that the laws of nature are constant’. To illustrate our point by a simple example, let us assume there is a ‘natural law’ (as Rothman et al. name it) for the association between E and D that reveals a relative risk of 1.0 if C = 1 (70% of the target population) and 5.0 if C = 0 (30% of the target population), that is, there is substantial effect modification of the association E*D by C (Figure 1). However, this effect modification has been completely unknown to the scientific community until now.

The investigator wants to study the association between E and D. To avoid representativeness (as Rothman et al. suggest), the investigator chooses to study only subjects with C = 1. If everything is done correctly (perfect validity and reliability), she would find a null result for E*D and would conclude that there is no association between E and D. She would stop performing further studies on this association. If she studies only subjects with C = 0, she would find a RR = 5.0 for E*D and would continue with further studies. Therefore, given these two extreme selection scenarios, there is only a 50% probability of detecting a new association that would promote further research and exploration of effect modification.

If the investigator had decided to select a more representative sample from the target population, she would have observed an average effect that had been weighted by the distribution of people across levels of C. This finding would prompt her to explore possible effect modification by C and other factors. Therefore, this approach would have a higher probability of discovering an association that could be further studied in detail. In the era of mega-cohorts and interest in gene-environment interaction, we think that at least some representativeness of the target population is helpful for making interesting discoveries as illustrated. We recently numerically illustrated this point in the context of the discussion of low response proportions at baseline examinations in cohort studies.3

To be sure, if the association between E and D would not be modified by a third factor C, we would fully agree with the idea that representativeness of study subjects is not a necessary requirement for drawing correct conclusions.
conclusions, as the association between E and D could be discovered in all subgroups selected for the study.

**References**


---

**Is representativeness the right question?**

From C Mary Schooling* and Heidi E Jones

City University of New York School of Public Health and Hunter College, New York, USA

*Corresponding author: E-mail: mschooli@hunter.cuny.edu

We agree completely that representative studies are immensely valuable for describing disease patterns, quantifying the burden of disease and generating risk stratification models. Given representativeness is time- and place-specific, these all need regular updates and more representative studies. For example, the SCORE (Systematic COronary Risk Evaluation) system for predicting fatal cardiovascular disease (CVD) uses the same risk factors across different models for high- and low-CVD-risk European countries but over time countries may, also, be promoted from high to low risk. Clearly, such risk prediction models are not scientific models that describe nature consistently across space and time, but they are immensely useful for service planning, targeting treatment and saving lives. Conversely, experimental studies, such as animal models and randomized controlled trials (RCTs), do not require representativeness to test scientific models.

On the other hand, whether observational epidemiological studies, representative or not, are useful for generating hypotheses or testing causal factors in scientific models is less clear. First, these represent the triumph of hope over experience. Second, as was pointed out over 20 years ago, nearly all possible hypotheses have already been generated. Third, some potentially relevant hypotheses may not be readily observed for conceptual or practical reasons. The current paradigm may exclude some hypotheses as impossible, making them imperceptible. Apart from well-known biases inherent in observational studies, causal factors may be invariant in commonly studied populations, expensive or difficult to measure, affected by preclinical disease or hidden within the (mis)classification of diseases by symptom rather than cause. Fourth, as a discipline we have not generally thought through the hierarchy of studies to refute a hypothesis. Our current methods, using the Bradford-Hill viewpoints as a touchstone, are much more focused on corroborating hypotheses, with an RCT as the pinnacle of corroboration. However, even something as simple as ‘field’ epidemiology may refute hypotheses. For example, the existence of populations with low birthweight and low rates of heart disease casts doubt on a major role of birthweight in heart disease.

Given these issues if we want to make progress in identifying causal processes in population health, assuming it is possible, rather than focusing on representativeness in studies used to generate or test (corroborate) hypotheses, it might be more useful to look for better ways to generate and screen plausible hypotheses, before we test them in suitable studies. Other methods of generating hypotheses about the drivers of population health are not obvious, but include using general mechanistic principles, starting with effective treatments and taking advantage of mechanistic insights from genetics or RCTs which include potential mediators. Not only do we need to move on from the debate about representativeness, we need to move onto some different questions.

**References**


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

3. Nohr EA, Olsen J. Commentary: Epidemiologists have debated representativeness for more than 40 years—has the time come to move on? *Int J Epidemiol* 2013;42:1016–17.