Our empirical paper was also misinterpreted. We found that selection into the internet-based birth cohort Nascita e INFanzia: gli Effetti dell’Ambiente (NINFEA) was affected by parity [1 + vs 0, odds ratio (OR) = 0.32], maternal education (graduate vs not graduate, OR = 2.30) and smoking during pregnancy (yes vs no, OR = 0.43), and very weakly affected (i.e. ORs between 1.1 and 1.3) by the other six analysed factors. Maternal smoking was relatively rare (prevalence of 8.9% in the population and of 3.7% in the cohort NINFEA) and we were thus left with only two potentially important (in terms of bias) factors: parity, and maternal education. We don’t agree that this is a very clear example of multiple factors having a strong effect on selection but, in any case, Ebrahim and Davey Smith did not consider that determinants of selection must also be important risk factors of the outcome in order to potentially cause bias. We considered two outcomes: caesarean section and low birthweight. Neither maternal education nor parity showed an adjusted association with these two outcomes larger than 2.0 or lower than 0.5 in the population. Moreover, they acted in opposite directions, thus resulting in little potential net confounding.

The main misrepresentation of our empirical study is more subtle, but also more fundamental. We aimed ‘to investigate how sample selection of the web-based NINFEA birth cohort may change the confounding patterns present in the source population’. Thus, we were not aiming to quantify the bias present in the internet-based birth cohort, as we did not believe that the population-based study represented a gold standard to compare with. Rather, we simply compared the two confounding patterns in the two populations. As also pointed out by Rothman and colleagues in their rebuttal, Ebrahim and Davey Smith seem to suggest that if a difference exists between a study conducted on a selected group and a study conducted in a general population, we should automatically believe in the latter. We don’t agree (for example we show, see Table 3 of our empirical paper, that a potential strong confounder in the general population can become a weak confounder in the selected sample).

It is surely not enough to hypothesize complex biases or to simply argue that bias seems unlikely. We believe that it is always important to quantify the likely directions and strengths of hypothesized biases, as we have attempted to do; of course, there may be different interpretations of such sensitivity analyses, but it is important that the analyses themselves are correctly reported.

References


Representativeness is not helpful in studying heterogeneity of effects across subgroups

From KJ Rothman, 1,2* EE Hatch 2 and JEJ Gallacher 3

1RTI Health Solutions, Research Triangle Park, NC, USA, 2Boston University School of Public Health, Boston, MA, USA and 3Cardiff University School of Medicine, Cardiff, UK

*Corresponding author. Research Triangle Institute, 3040 Cornwallis Road, P.O. Box 12194, Research Triangle Park, NC 27709-2194, USA. E-mail: krothman@rti.org

In response to our commentary that representativeness best be avoided, Stang and Jöckel argue that representativeness is not needed in studying uniform effects, but that it is helpful when studying effects that vary across subgroups. In our original papers, we did address this point: ‘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’ and ‘When studying...
effects across a range of a variable such as age, representativeness is not the most effective way to do so.\(^3\)

Nevertheless, we are grateful to Stang and Jöckel for an opportunity to illustrate our point in greater detail. Suppose one studies an exposure-disease relation in a population with three ethnic subgroups, A, B and C, which comprise 60%, 30% and 10% of the population, respectively. Further suppose that the association under study is heterogeneous, varying across the three groups. Representativeness would lead to three subgroups in our study population of greatly differing size, in the ratio of 6:3:1. Study participants from ethnic group C will constitute only a sliver of the total study population. This representative distribution will be highly inefficient for comparing the magnitude of the association across the three groups. Much better efficiency would be achieved by enrolling a third of the study population from each ethnic group, giving equal information for each of A, B and C. The key issue to addressing the scientific goal is not obtaining a representative study population, but rather including sufficient numbers of subjects across a range of the variable that modifies the association. If one suspects heterogeneity, one should design a study deliberately to study how the modifying variable changes the association, which will usually mean choosing study participants so that numbers across the compared groups are balanced. But balance across groups is a different concept from representativeness, and is much more efficient than representativeness in achieving the study goal.

Schooling and Jones\(^4\) assert that representativeness is not a requirement for scientific studies, agreeing with our primary thesis. We admire their attempt to reconcile epidemiological progress with Popper’s paradigm of conjecture and refutation,\(^5\) and we agree with their argument that focusing on representativeness in scientific studies detracts from more fruitful activities such as hypothesis formulation, notwithstanding the advent of Cole’s tongue-in-cheek hypothesis-generating machine.\(^6\) Nonetheless, we depart from them when it comes to their pessimism about the scientific value of nonexperimental epidemiology.

References