Defining representativeness of study samples in medical and population health research

Jacqueline E Rudolph [iD] , Yongqi Zhong, Priya Duggal, Shruti H Mehta, Bryan Lau

Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA
Correspondence to: Dr Jacqueline E Rudolph, Department of Epidemiology, Johns Hopkins University, Baltimore, MD 21205, USA; jacqueline.rudolph@jhu.edu

ABSTRACT

Medical and population health science researchers frequently make ambiguous statements about whether they believe their study sample or results are representative of some (implicit or explicit) target population. This article provides a comprehensive definition of representativeness, with the goal of capturing the different ways in which a study can be representative of a target population. It is proposed that a study is representative if the estimate obtained in the study sample is generalisable to the target population (owing to representative sampling, estimation of stratum specific effects, or quantitative methods to generalise or transport estimates) or the interpretation of the results is generalisable to the target population (based on fundamental scientific premises and substantive background knowledge). This definition is explored in the context of four covid-19 studies, ranging from laboratory science to descriptive epidemiology. All statements regarding representativeness should make clear the way in which the study results generalise, the target population the results are being generalised to, and the assumptions that must hold for that generalisation to be scientifically or statistically justifiable.

Introduction

It is common if not a requirement for medical and population health science researchers to consider the inferences from a study beyond the context of their analysis. Accordingly, many papers mention whether their study sample is representative of, or study results generalise to, some implicit or explicit target population; others refer to a lack of generalisability or representativeness as a limitation. Despite being frequently discussed and debated,[1–4] in common practice, the meaning of representativeness remains ambiguous. Here, we propose a comprehensive definition of representativeness and discuss it in the context of different study designs. We presume no bias in the study's results; in any real world study, bias will need to be weighed alongside whether the sample is representative or its results are generalisable or applicable to a target population.[5]

What is representativeness?

The ambiguity in meaning arises in part because the word "representative" has a broader meaning in English and a more technical definition, and the definition being used is not always clear. In a 2013 series of commentaries on representativeness,[1–4] the concept was defined as occurring when the study sample is a simple random sample of the target population (ie, the sample that arises through representative sampling). A second definition is that the study sample and the results obtained merely resemble what would be expected in the target population, perhaps based on a similarity in personal characteristics.[6] The first definition is more precise and implies a high standard for study design, while the second encompasses a variety of possible interpretations.

Here, we bridge these two uses of the word "representative" and attempt to concretise the second, broader definition. We define a study sample to be representative of a well defined target population if the results estimated in that sample are generalisable to the target population. We consider two ways in which study results can generalise to the target population: in estimate and in interpretation. Box 1 lists a summary of key terms used in this article and figures 1–2 show examples applying the definition of representativeness.

To help clarify these definitions, we use as an example a randomised controlled trial that was conducted to measure the efficacy of molnupiravir for treatment of covid-19.[7] Among unvaccinated adults with mild to moderate covid-19 who were not in the hospital, researchers found that the risk of hospital admission or death at 29 days among participants randomised to molnupiravir was 6.8%, compared with 9.7% of participants randomised to placebo. They concluded that treatment with molnupiravir within five days of infection reduced the risk of hospital admission or death. In boxes 2–4, we explore the definitions in the context of other study designs.[8–10] As with the trial, we use the question, study design, and sample description of these publications to construct a theoretical example. We do not delve into the specific details of the study or comment on whether the researchers fully achieved what we discuss.

KEY MESSAGES

⇒ Researchers frequently refer to whether their sample is or is not representative without clarifying whether they mean that their sample is a simple random sample of their target population or that the results from their sample are merely reflective of what would be seen in the target population

⇒ This article provides a comprehensive definition of what it means for a study to be representative, and examines this definition in the context of examples with different study designs

⇒ When publishing research, researchers should critically assess whether a study sample is representative of a clearly defined target population, by carefully considering the manner in which they think the results generalise to the target population and the assumptions underlying that hypothesis

BOX 1 | GLOSSARY

⇒ **Effect measure modifiers:** variables that influence (ie, weaken or strengthen) the relation between the treatment and outcome

⇒ **Estimate:** numerical result (eg, mean, risk, risk difference, risk ratio, odds ratio) obtained in the study sample

⇒ **Generalisable:** findings in the study sample study apply to an overlapping target population (ie, study sample is at least a partial subset of the target population)

⇒ **Interpretation:** knowledge or information learnt from the numerical estimate, such as the direction of effect or other study conclusions

⇒ **Key covariates:** variables that affect the outcome, which might be effect measure modifier on some scale (eg, risk difference, risk ratio, odds ratio) and must be considered if generalising the estimate to a target population

⇒ **Representative:** a study sample is representative of a well defined target population if either the estimate obtained in that sample or the interpretation of the results in that sample are generalisable to the target population

⇒ **Target population:** population to which the researcher seeks to make inference

⇒ **Transportable:** findings in the study sample study apply to a non-overlapping target population (ie, study sample is not a subset of the target population)

Generalisable in estimate

A sample is representative if its results are generalisable in estimate. For a given estimand (eg, risk difference, odds ratio, population mean), the estimate obtained in the study sample is the same within a margin of error as what would be estimated in the target population. In the molnupiravir randomised controlled trial,[7] we might hypothesise that the risk difference comparing molnupiravir with placebo estimated in the trial is the same risk difference as would be estimated in the target population of all adults with recent SARS-CoV-2 infection who were not in the hospital. Generalising the estimate obtained in a given study sample might be considered the primary goal when intending to quantitatively inform policy interventions or when obtaining effect estimates in the target population is impossible or infeasible.[11]

Generalisability in estimate can be achieved if the distributions of key covariates are the same as in the target population, as would occur in expectation with random sampling. Thus, generalising the estimate aligns closely with the definition of representativeness based on representative sampling. These key covariates are those that affect the variable under study (eg, hospital admission or death) and thus are potential effect measure modifiers of the effect of a treatment on that variable. By effect measure modifiers, we mean variables where the effect of the treatment differs by levels of that variable on some scale (eg, risk difference, risk ratio, odds ratio). In our example, age might be an effect measure modifier because the effect of molnupiravir on hospital admission or death (as quantified by the risk difference) might differ across ages.

More generally, even if the distribution of the key covariates differs between the sample and target population, the sample might still be representative within stratums of the key covariates, such that the stratum specific estimates (eg, risk difference within age categories) can be generalised from the sample to the target population. While this generalisation requires that all the key covariates be measured, the proportion of the sample in the covariate stratums need not exactly match the proportion who fall into that subgroup in the target population.

If we apply this definition of generalisability in estimate to the molnupiravir trial example, suppose our target population is all individuals recently infected with SARS-CoV-2 who were not in the hospital. In this case, we would not be able to generalise the trial's estimate to the entire target population, even if we could control for post-randomisation factors such as non-adherence, because the trial sample did not include vaccinated individuals. However, the target population does include vaccinated individuals, and it is reasonable to assume that the effect of molnupiravir on disease progression to hospital admission or death would vary by vaccination status. On the other hand, we might be able to generalise our results within the stratum of unvaccinated individuals, provided all other effect measure modifiers were similar. In this case, we would say that the sample is representative within that stratum.

Generalisable in interpretation

A sample is representative if its results are generalisable in interpretation. While the estimates obtained in the sample are not quantitatively the same within a margin of error as those that would be estimated in the target population, we can hypothesise, based on background knowledge, that the interpretation (which could be the direction of effect, general inference from the results, or knowledge gained from an experiment) would remain the same.[12] For example, we might hypothesise that molnupiravir is generally protective against hospital admission or death from covid-19, even in samples other than the study sample. Generalising the interpretation aligns with the broad definition of representative, which states that a study sample resembles what would be expected in the target population.

We often generalise the interpretation of our own results to external populations, and any study that generalises in estimate will also generalise in interpretation. The primary goal in studies examining fundamental laws of nature or asking research questions that are relatively independent of historical and
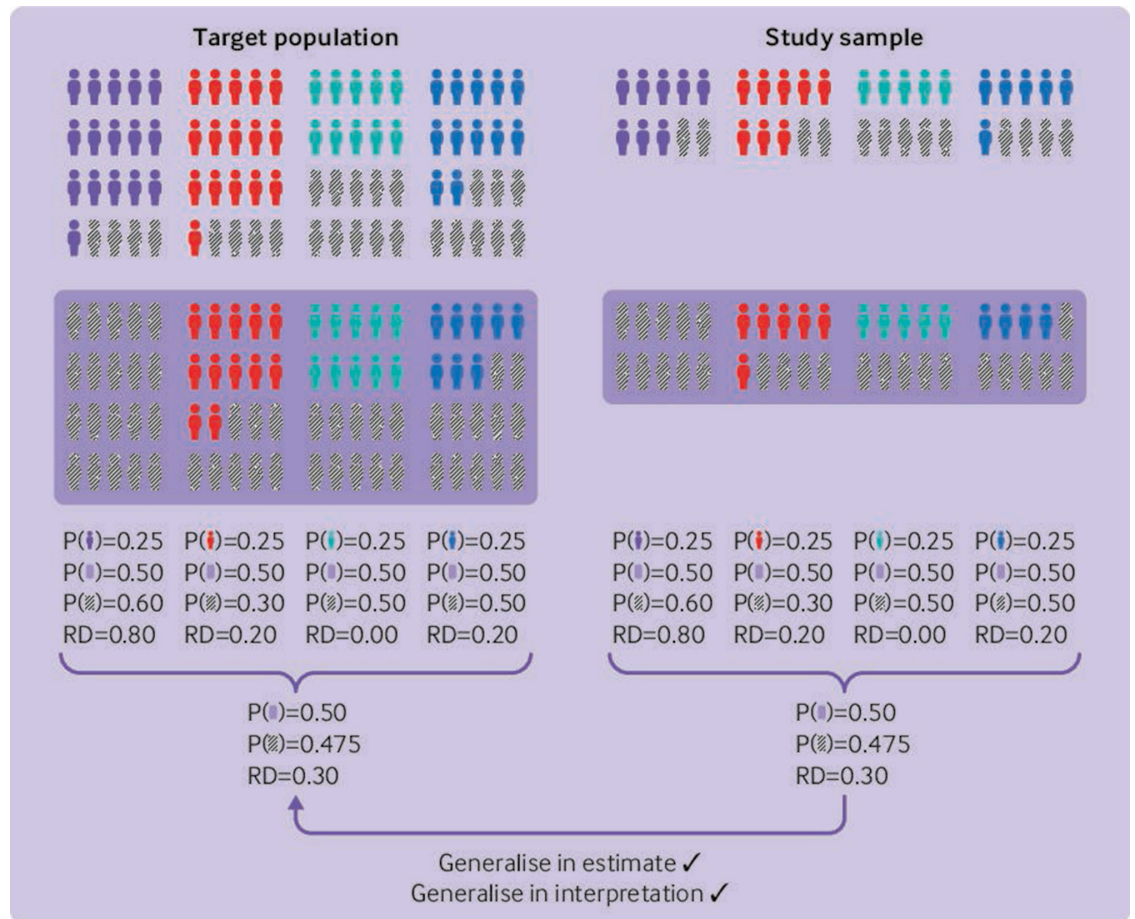
Figure 1 | Example where a study sample (which is a simple random sample of the target population) is representative because its results generalise in interpretation and in estimate. Shaded box=treatment group; hashed lines=outcome group. Colours represent different levels of an effect measure modifier on the risk difference scale, which did not affect selection into the study sample

environmental context are to generalise the interpretation, rather than the estimate. Generalising the interpretation should be done cautiously, however, because it is based on hypotheses that the mechanisms or biological processes under investigation in the study sample are (at least approximately) identical to those that would be seen in the target population.

If we apply this definition of generalisability in interpretation to the molnupiravir trial example,[7] it might be reasonable to hypothesise that molnupiravir would have a beneficial impact if given to those individuals infected with SARS-CoV-2, even beyond the enrolled sample of participants with moderate illness who were not in the hospital. We might base this hypothesis on our understanding of the drug's biological mechanism and the validity of a properly conducted, double blind, placebo controlled trial. In this example, we generalise the interpretation despite the fact that the risk difference comparing molnupiravir with placebo estimated in the trial would differ from the risk difference estimated in the target population (again assuming that the target

population is all individuals with recent SARS-CoV-2 infection who were not in the hospital).

Discussion

In summary, we consider a sample to be representative of a target population if its results can be generalised to that target population either in estimate or in interpretation. Any statements made regarding the representativeness of the study need to make this further qualification. Is it the estimate obtained or the interpretation of the results that are generalisable to the target population? Researchers should also do what they can to safeguard their results from being applied incorrectly. Even in studies with a strong scientific rationale for generalising the interpretation of results to the target population, researchers might need to mention that the estimate obtained in the sample should not be naively generalised to the target population.

Stating which form of representativeness was the goal of the study might also be useful. In the example of the molnupiravir randomised controlled trial,[7] generalising the interpretation regarding drug
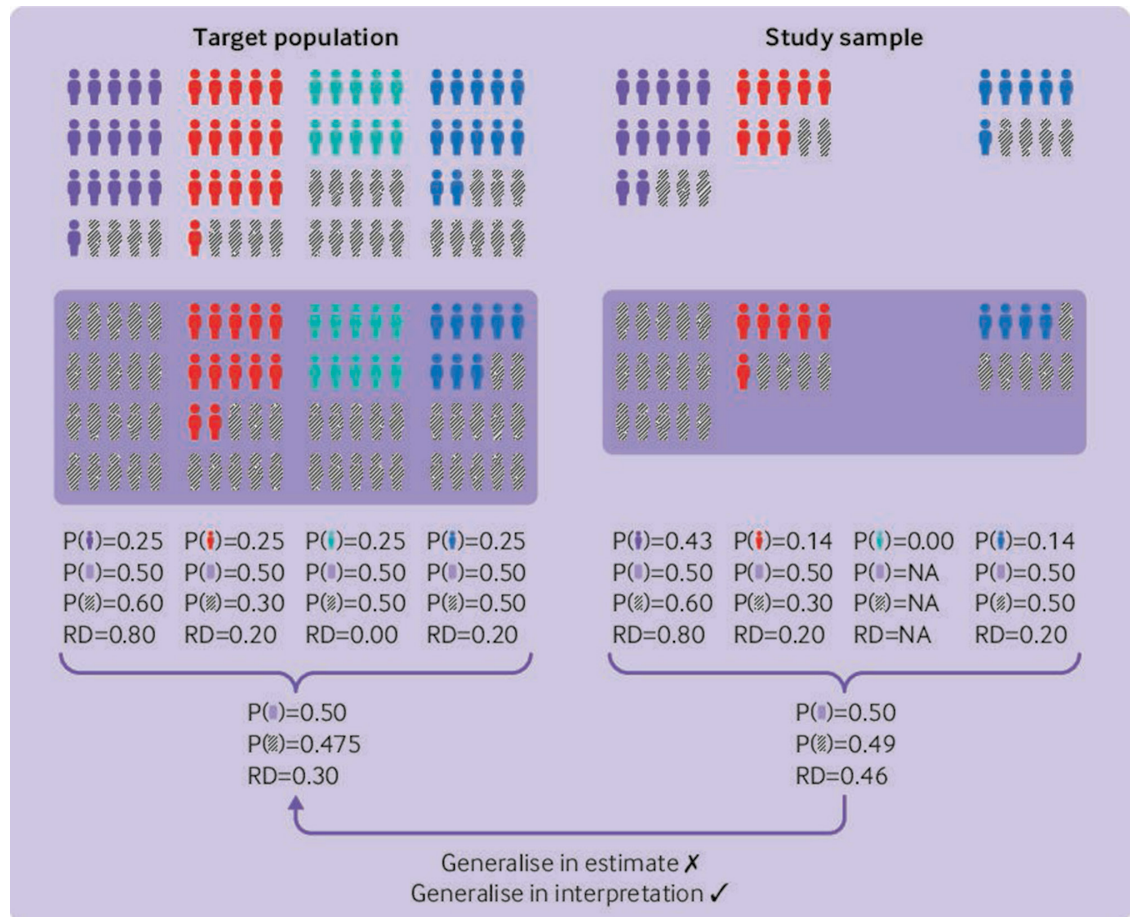
Figure 2 | Example where a study sample (which is a convenience sample of the target population) is representative because its results generalise in interpretation even though they do not generalise in estimate. Shaded box=treatment group; hashed lines=outcome group. Colours represent different levels of an effect measure modifier on the risk difference scale, which affected selection into the study sample

efficacy to the target population might have been the primary goal. Many trials have this same goal, because the investigators often over sample individuals at high risk for the outcome in order to increase the power of the study. (Even so, clinical trials have received some criticism that they rarely represent a more general target population.[5]) If it was possible, generalising the estimate to the target population would be useful for predicting how molnupiravir would perform in practice but might not be immediately required for the study results to be meaningful. Further studies would likely need to be conducted to generalise the interpretation to other target populations, such as children recently infected with SARS-CoV2.

Several points relate to defining representativeness and are worth discussing. Firstly, irrespective of the way in which a sample is representative, the target population must be clearly defined. Stating that a sample is representative is meaningless unless researchers specify what population it represents or its results are being applied to.[5] As an example, we showed how specifying different target populations (all individuals *v* all unvaccinated individuals

who were not in the hospital) for the molnupiravir randomised controlled trial had different implications for whether the results were generalisable in estimate.

Secondly, researchers must be clear about the assumptions required for generalising to the target population. When generalising the estimate, these assumptions might be made based on knowledge of whether the study was designed using a simple random sample or whether stratification by relevant key covariates is possible. When generalising the interpretation, the assumptions might be made based on a knowledge of basic scientific premises or the validity of a related animal model. If researchers attempted to generalise the interpretation but the scientific principles underlying that generalisation did not hold (eg, the validity of the animal model for describing human physiology), then the assumptions would be violated, and the inferences in the study would not be representative. In either case, the way to truly test whether the assumptions held would be to estimate the effect of interest in the target population. While we often generalise in estimate because designing a study in the target population would

BOX 2 | DEFINING REPRESENTATIVENESS IN LABORATORY SCIENCE STUDIES

At the start of the covid-19 pandemic, no antiviral drugs for infection or disease had been approved. To assess whether antiviral drug molnupiravir was effective for treating covid-19, researchers in one animal model study gave molnupiravir to mice with human lung tissue before and after infection with SARS-CoV-2, using doses scaled from appropriate human levels to the mouse model.[8] They found that a 2 day course of treatment, started 24 hours after infection, significantly reduced SARS-CoV-2 viraemia in lung tissue.

TARGET POPULATION

All humans with recent SARS-CoV-2 infection.

GENERALISABILITY OF INTERPRETATION

Yes. We likely can hypothesise that the beneficial effect of molnupiravir observed in the mice would be observed in humans, based on the validity of the human lung tissue model and on the observation of a similar pathological response to covid-19 in the lung tissue of the mice as has been seen in the lung tissue of patients with covid-19.

GENERALISABILITY OF ESTIMATE

No. While the study used human lung tissue in mice and used an appropriately scaled dose, the lung tissue was otherwise isolated from human biology, and mouse immune responses differ from those seen in humans in a manner that would be very difficult to quantify.

OVERVIEW

In this animal model study, generalising the interpretation of the results was the primary goal. Generalisation of the estimate was not relevant, as the drug would be tested further in human studies. When it comes to animal model studies (or cell line studies), we need to recognise that an underlying assumption is that the strength of the unaccounted-for effect measure modifiers (which are likely unknown) and the difference in the distribution of the effect measure modifiers between the study sample and human target population is not large enough to change the inference being made. This strong assumption is why animal studies are followed up by clinical trials to ensure that interpretation does indeed generalise and to obtain a quantifiable estimate of the effect in humans.

BOX 3 | DEFINING REPRESENTATIVENESS IN OBSERVATIONAL STUDIES

Researchers used testing, hospital, and vaccine registry databases to build an observational cohort of vaccinated adults living in New York, with age matched unvaccinated controls.[9] Their goal was to assess the effectiveness of covid-19 vaccines for preventing SARS-CoV-2 infection and covid-19 related hospital admission in the general population. They found that vaccine effectiveness for preventing infection was highest in the week of 1 May 2021 (93.4%) (when prevalence of the delta variant was negligible), but that effectiveness declined as the delta variant became more prevalent, with a low of 73.5% in the week of 10 July 2021. By contrast, the effectiveness for preventing hospital admission did not wane during this same calendar period. The researchers concluded that their findings were evidence in support of booster vaccines.

TARGET POPULATION

Adult residents of New York state.

GENERALISABILITY OF INTERPRETATION

Yes. We have little reason to suspect that vaccines would not be effective against covid-19 and hospital admission or that we would observe different trends in vaccine effectiveness over time among New York residents who were not included in this study (or residents of other US states).

GENERALISABILITY OF ESTIMATE

Additional data needed. The registry based study included a wide range of ages and the different vaccine types (Pfizer, Moderna, and Johnson & Johnson). The paper did not report the distribution of comorbid conditions such as asthma, which could be potential effect measure modifiers. With such information, researchers may be able to determine whether the estimate could be generalised to the broader New York population.

OVERVIEW

Here, generalising the interpretation and the estimate are both important. Generalising the estimate to the target population requires more effort both from the researchers designing the study and from those analysing the data but could be incredibly useful for informing covid-19 prevention efforts in New York. If we wished to generalise to target populations beyond New York (eg, the entire US), we would need to make assumptions about whether there are effect measure modifiers that differ between New York and the entire US and whether we have them measured.

not be feasible, we generally consider such a study necessary to prove hypotheses regarding generalisation of interpretation, especially when the sample is highly removed from the target population (eg, cell line v human population).

Thirdly, a natural extension of generalising the (overall or stratum specific) estimate to a target population are methods to estimate the overall mean of an outcome or the average effect of a treatment on an outcome (rather than a stratum specific estimate)

<div style="border: 1px solid;">

BOX 4 | DEFINING REPRESENTATIVENESS IN DESCRIPTIVE STUDIES

Researchers sought to capture the burden of covid-19 among people who inject drugs in the San Diego-Tijuana area.[10] Participants from both cities were recruited using street outreach and mobile vans. Blood samples and nasal swabs were collected to test for the presence of SARS-CoV-2 antibodies and RNA. None of the 485 participants had detectable SARS-CoV-2 RNA, but 140 (36.3%) were seropositive based on the presence of antibodies. This proportion was larger than the prevalence reported in the general population for either city. No trends were seen in prevalence of antibodies to SARS-CoV-2 over the study period (October 2020-June 2021).

TARGET POPULATION

People who use drugs by injection in the San Diego-Tijuana region.

GENERALISABILITY OF INTERPRETATION

Yes. It is reasonable that the target population has a higher prevalence of SARS-CoV-2 than the general population, even beyond the time frame and sample studied.

GENERALISABILITY OF ESTIMATE

Perhaps. Under an appropriate sampling and recruitment strategy, the 36.3% prevalence of SARS-CoV-2 could be generalised to the full sample of the target population, at least during the time frame examined. We would be unable to generalise the estimate to other points in the pandemic, with different SARS-CoV-2 strains and levels of community exchange.

OVERVIEW

Just as in the observational study, generalising both the estimate and the interpretation are important for assessing the relevance of this study. We note here that the target population selected was much narrower than those of the previous studies, but this reflects the research and public health goals of the study. The researchers likely could not make statements regarding representativeness to broader target populations (eg, all people who inject drugs in the US) without further evidence.

</div>

in the target population.[5 13 14] While the study sample might not be representative of the target population as observed, it could be made representative by using methods for generalisability or transportability, such as weighting or standardization to control for the key covariates or effect measure modifiers that differ between the samples.[15] These approaches require measuring and accounting for all relevant key covariates, meeting certain identifiability conditions, and often making model specification assumptions.[13 14]

Even further, any study that is representative in interpretation could theoretically be made representative in estimate if all relevant effect measure modifiers were measured and accounted for; however, that is not always possible when the study sample is distant from the target population (eg, laboratory mice to humans).

Fourthly, the concepts of representativeness and generalisability discussed above also relate to the term "applicability" used in certain risk-of-bias tools, such as the PROBAST and QUADAS.[16 17] All concepts centre on the idea that it is important to assess a study and its results in terms of how well they can be related to some target population. While we discussed causal and descriptive studies in this article, the two tools mentioned apply this concept to predictive and diagnostic studies.

Finally, one question that has been raised is whether generalising the interpretation or the estimate is intrinsically more important for health research and for science broadly. It could be argued that generalising the interpretation is the primary aim of scientific inference and thus should be our goal in most studies.[1] The underlying premise is that the goal of science is the discovery of universal knowledge about nature that will hold true in most instances. If we view health research from this viewpoint, then generalising the interpretation is what matters. By contrast, generalisation of the estimate can never be universal. The estimate obtained in a particular study sample will always be tied to a specific scientific or public health question, and the study design and will vary based on the distribution of key covariates across time and populations. However, to inform policies and interventions in the real world, we must be able to predict health outcomes in human populations beyond those we studied. Therefore, generalisation of the estimate (whether obtained via study design or analytical methods) is an important goal. A further argument could be that these endeavours of statistical inference are just as informative for science as the inferences above. Science can be about discovering laws of nature; it can also seek to understand particular facets of nature. For some areas of health research, such as epidemiology and other population health sciences, the facet of nature under study is disease as it occurs in humans at a population level, and true understanding of the disease under study will be contextualised by time, place, history, and social environment. Consideration for how these factors have changed from the original setting to some new time or target population and how these changes might affect the estimate obtained is critical.

While such theoretical debates are important, our comprehensive definition of representativeness does not treat either generalisation of estimate or interpretation as inherently more relevant. That evaluation largely depends on the research question and study

design at hand. Health researchers both develop the universal knowledge related to the health of populations and investigate how that knowledge can be applied to improve the health of populations, and the two ends of the research spectrum are fundamentally linked. What is important, then, is that researchers are clear on the manner in which their results can be applied to the target population when they say their study is representative and the assumptions underlying that statement.

Conclusions

We have established the idea that a study sample can be representative of a target population if one of the following is true: the estimate obtained in the study sample is generalisable to the target population or the interpretation of the study results is generalisable to the target population. Whether a study sample can be representative of a target population through the first definition depends on the study design or whether the variables affecting the outcome (which could be effect measure modifiers of the effect of interest) have been measured. On the other hand, even in the absence of simple random sampling or measurement of all key covariates, we can say that the study is representative in terms of its interpretation, direction of effect, or inference, because this requires less stringent assumptions than generalising the study estimate.[12] The example studies provided give guidance on how one might determine whether the study sample from different types of research is representative and whether, for the specific research question, generalising the estimate or the interpretation was the priority.

ORCID iD

Jacqueline E Rudolph http://orcid.org/0000-0001-7177-1847

REFERENCES

1 Rothman KJ, Gallacher JEJ, Hatch EE. Why representativeness should be avoided. Int J Epidemiol 2013;42:1012–4. 10.1093/ije/dys223
2 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;42:1016–7. 10.1093/ije/dyt102
3 Elwood JM. Commentary: on representativeness. Int J Epidemiol 2013:1014–5. 10.1093/ije/dyt101
4 Richiardi L, Pizzi C, Pearce N. Commentary: representativeness is usually not necessary and often should be avoided. Int J Epidemiol 2013;42:1018–22. 10.1093/ije/dyt103
5 Westreich D, Edwards JK, Lesko CR, et al. Target validity and the hierarchy of study designs. Am J Epidemiol 2019;188:438–43. 10.1093/aje/kwy228
6 Last JM. Representative sample. In: A Dictionary of Epidemiology. New York: Oxford University Press, 1983.
7 Jayk Bernal A, Gomes da Silva MM, Musungaie DB, et al. Molnupiravir for oral treatment of covid-19 in nonhospitalized patients. N Engl J Med 2022;386:509–20. 10.1056/NEJMoa2116044
8 Wahl A, Gralinski LE, Johnson CE, et al. SARS-cov-2 infection is effectively treated and prevented by EIDD-2801. Nature 2021;591:451–7. 10.1038/s41586-021-03312-w
9 Rosenberg ES, Dorabawila V, Easton D, et al. Covid-19 vaccine effectiveness in new york state. N Engl J Med 2022;386:116–27. 10.1056/NEJMoa2116063
10 Strathdee SA, Abramovitz D, Harvey-Vera A, et al. Prevalence and correlates of SARS-cov-2 seropositivity among people who inject drugs in the san diego-tijuana border region. PLOS ONE 2021;16:e0260286. 10.1371/journal.pone.0260286
11 Rudolph JE, Cole SR, Eron JJ, et al. Estimating human immunodeficiency virus (HIV) prevention effects in low-incidence settings. Epidemiology 2019;30:358–64. 10.1097/EDE.0000000000000966
12 Egami N, Hartman E. Elements of external validity: framework, design, and analysis. Am Polit Sci Rev 2022:1–19. 10.1017/S0003055422000880
13 Lesko CR, Buchanan AL, Westreich D, et al. Generalizing study results: a potential outcomes perspective. Epidemiology 2017;28:553–61. 10.1097/EDE.0000000000000664
14 Degtiar I, Rose S. A review of generalizability and transportability. Annu Rev Stat Appl 2023;10:501–24. 10.1146/annurev-statistics-042522-103837
15 Westreich D, Edwards JK, Lesko CR, et al. Transportability of trial results using inverse odds of sampling weights. Am J Epidemiol 2017;186:1010–4. 10.1093/aje/kwx164
16 Wolff RF, Moons KGM, Riley RD, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann Intern Med 2019;170:51–8. 10.7326/M18-1376
17 Bristol U of. QUADAS-2 [Internet]. University of Bristol. 2022. Available: https://www.bristol.ac.uk/population-health-sciences/projects/quadas/quadas-2/