European Journal of Epidemiology (2019) 34:723–724
https://doi.org/10.1007/s10654-019-00532-3

COMMENTARY

New methods for generalizability and transportability: the new norm

Sunni L. Mumford[1] · **Enrique F. Schisterman**[1]

Received: 29 May 2019 / Accepted: 1 June 2019 / Published online: 7 June 2019

Generalizability is an ever-present concern in epidemiology. Though randomized trials are often considered the gold standard, as Weiss describes there are often scenarios where results from randomized trials are not more broadly applicable and observational studies are needed to inform associations in populations not represented in randomized trials [1]. We recently completed a randomized trial that highlights the importance of generalizability and of considering the target population from the outset. The Effects of Aspirin in Gestation and Reproduction (EAGeR) trial was designed to evaluate the effect of low dose aspirin on pregnancy loss [2]. It is hypothesized that aspirin may improve blood flow in the uterus, thus facilitating implantation, and that this may be particularly beneficial for women who have recently had a pregnancy loss. Aspirin is well known for its anti-inflammatory properties, and as it is a low cost and widely available intervention, if it could be broadly applied could have a tremendous public health impact.

With this in mind, we targeted two populations and enrolled women who were actively trying to conceive in either an 'original' or 'expanded' stratum. The 'original' stratum could also be thought of as the 'biologically-based inclusion criteria' as it included women with a history of a single prior pregnancy loss at less than 20 weeks' gestation during the previous year [2, 3]. Women with a single recent pregnancy loss might have an endometrium that is in the healing process and may be more likely to benefit from the positive effect of low dose aspirin on reducing inflammation and enhancing blood blow. Further, women with a recent loss may be more motivated to participate in a trial. However, knowing that recruitment for this specific population would be challenging, and that findings from this trial would likely be applied in practice with small deviations from the original inclusion criteria (e.g., 1 or 2 prior losses or losses at later gestational ages), we also enrolled women in an 'expanded' stratum. In this population the inclusion criteria were extended to women with either one or two prior pregnancy losses at any time in the past, including those that occurred at more than 20 weeks' gestation. Women with 3 or more losses were not included in the trial as they would meet the criteria for recurrent pregnancy loss, which likely would not benefit from aspirin therapy in the same way. This 'expanded' population is what we would consider the more pragmatic trial as these inclusion criteria are more open and more akin to how the intervention would be applied in practice.

By design, the EAGeR trial allowed us to estimate effects in two seemingly very similar, yet distinct populations. Indeed, the effects of aspirin were different by strata. When using the biologically based inclusion criteria we saw a beneficial effect of low dose aspirin use on live birth rates (RD 9.20%, 95% CI: 0.51–17.89), however, among the broader population of women with pregnancy losses, the effect was attenuated (RD 1.71%, 95% CI: −6.37–9.79) [3]. This trial demonstrated the biological proof of concept for low dose aspirin to improve live birth rates among women with a single recent loss, though also showed that the intervention may not be beneficial outside of this select population. It further shows that even seemingly small changes to the population may have important downstream implications.

It may not always be possible to include multiple strata in the design of a randomized trial to address these issues. In such cases one must consider the scenarios raised by Weiss [1], in combination with the innovation of recent methods to address generalizability. It is also important to recognize the nuances to the language in this area and that "generalizability" has a very specific meaning in epidemiology and cannot be applied more broadly. Generalizability refers to problems arising when the study sample is a subset of the target population, whereas transportability refers to problems arising when the study sample is not a subset of the target population [4]. This distinction has important implications when finding the appropriate method to generalize or transport

✉ Sunni L. Mumford
mumfords@mail.nih.gov

[1] Epidemiology Branch, Division of Intramural Population Health Research, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD, USA

findings to other populations. There are several potential methods to improve the generalizability and transportability of randomized trials including post-stratification, meta-analysis, research synthesis approaches, confidence profile method, and reweighting approaches including propensity scores (discussed in Stuart et al. [5]). Inverse probability weights have been described for quantitative generalization of trial results [6–8], as well as a transport formula and inverse odds of sampling weights which work to transport effect estimates to target populations [4, 9]. These methods typically require the assumptions of conditional exchangeability, positivity, no measurement error, treatment variation irrelevance, and no misspecification of the model [4]. However, while some are uncomfortable with the assumptions required for these methods, it should be noted that these are common assumptions and these methods are importantly making these assumptions explicit. With modern methods, the issues described by Weiss are perhaps less of a problem even in restricted trials [1].

Moreover, and perhaps most importantly, safety is paramount. Even when multiple strata are included to address transportability in terms of efficacy, many randomized trials are too small to adequately evaluate adverse events and safety. Indeed, no randomized trial is ever powered to look at safety. Large cohort studies are thus needed to get a full picture of the safety profile of interventions. Though current regulatory bodies, such as the FDA, and other decision making organizations are currently working to synthesize the available evidence, the new methods for generalizability and transportability are likely needed. In reality, with the development of these new methods, the transportability of our findings to other cohorts is possible and should become standard practice.

Given the innovation of methods to address generalizability and transportability, these issues should be part of the description of any trial. Failing to recognize the explosion of methods to extend results of randomized trials may lead us to erroneous conclusions, which in the context of public health decision making could be disastrous. Clinicians and policymakers rely on findings from randomized trials and observational studies to make evidence-based decisions. Therefore, it is more important than ever to be explicit regarding the target population to further the usefulness of interventions to improve public health.

Funding This work was supported by the Intramural Research Program of the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health.

References

1. Weiss NS. Generalizing from the results of randomized studies of treatment: can non-randomized studies be of help? Eur J Epidemiol. 2019. https://doi.org/10.1007/s10654-019-00516-3.
2. Schisterman EF, Silver RM, Perkins NJ, Mumford SL, Whitcomb BW, Stanford JB, Lesher LL, Faraggi D, Wactawski-Wende J, Browne RW, Townsend JM, White M, Lynch AM, Galai N. A randomised trial to evaluate the effects of low-dose aspirin in gestation and reproduction: design and baseline characteristics. Paediatr Perinat Epidemiol. 2013;27:598–609.
3. Schisterman EF, Silver RM, Lesher LL, Faraggi D, Wactawski-Wende J, Townsend JM, Lynch AM, Perkins NJ, Mumford SL, Galai N. Preconception low-dose aspirin and pregnancy outcomes: results from the EAGeR randomised trial. Lancet. 2014;384:29–36.
4. Westreich D, Edwards JK, Lesko CR, Stuart E, Cole SR. Transportability of trial results using inverse odds of sampling weights. Am J Epidemiol. 2017;186(8):1010–4.
5. Stuart EA, Bradshaw CP, Leaf PJ. Assessing the generalizability of randomized trial results to target populations. Prev Sci. 2015;16:475–85.
6. Cole SR, Stuart EA. Generalizing evidence from randomized clinical trials to target populations: the ACTG 320 trial. Am J Epidemiol. 2010;172(1):107–15.
7. Lesko CR, Buchanan AL, Westreich D, Edwards JK, Hudgens MG, Cole SR. Generalizing study results: a potential outcomes perspective. Epidemiology. 2017;28:553–61.
8. Buchanan AL, Hudgens MG, Cole SR, Mollan KR, Sax PE, Daar ES, Adimora AA, Eron JJ, Mugavero MJ. Generalizing evidence from randomized trials using inverse probability of sampling weights. J R Stat Soc Ser A Stat Soc. 2018;181(4):1193–209.
9. Bareinboim E, Pearl J. A general algorithm for deciding transportability of experimental results. J Causal Inference. 2013;1(1):107–34.