Editorial / Editöryal Yorum

Generalizability and transportability of research findings: Randomized trials vs observational studies

Araştırma bulgularının genelleştirilmesi ve taşınabilirliği: Randomize denemeler ve gözlemsel çalışmalar

Oğuzhan Birdal, M.D.,[1] İbrahim Halil Tanboğa, M.D., PhD[2,3]

[1]Department of Cardiology, Atatürk University School of Medicine, Erzurum, Turkey
[2]Department of Biostatistics, Nişantaşı University School of Medicine, İstanbul, Turkey
[3]Department of Cardiology, Nişantaşı University School of Medicine, İstanbul, Turkey

Randomized clinical trials (RCTs) are considered as the gold standard for evaluating the effect of a treatment or an intervention. Some basic definitions about RCTs are shown in Table 1 and Figure 1. RCTs are designed to obtain absolute treatment effect in continuous outcome and relative treatment effect in binary or time-to-event outcomes.[1] Through this editorial, RCT examples that evaluated binary or time-to-event outcomes will be discussed. Many researchers claim that RCTs have strong internal validity and weak external validity, whereas observational studies have weak internal validity and strong external validity. They claim that the trial samples in RCTs are not representative of the target population and the participants of observational studies are more representative of it; thus, their external validity (generalizability and transportability of the results) is stronger than that of RCTs.[2,3]

Altın et al.[4] have conducted a prospective, multicenter, and postmarketing observational study for evaluating the real-life safety and effectiveness of dabigatran etexilat (D-SPIRIT). Investigators included 326 patients with atrial fibrillation who had been using dabigatran etexilat for at least 6 months before enrollment and followed them for treatment effectiveness and safety for 2 years. They reported the rate of embolic complications was 1.26% per year, major bleeding was 2.20% per year, and mortality was 0.94% per year. Interestingly, all reported event frequencies were lower than the event frequencies in the RELY trial.[5] Therefore, the investigators concluded that the results of RELY trial are validated in these observational data. This study is important because it is the first study of its kind published in our country. It also carries valuable information about how to administer doses in clinical practice. Although the investigators wished to evaluate the effectiveness and safety of dabigatran, the current study design (lack of a control group) precludes this.

Abbreviation:
RCTs *Randomized clinical trials*

Criticisms of RCT in terms of their weak external validity are based on the assumption that the individuals participating in the trial should be random samples from the target population and that they should be a representative sample of this target population, but many authors argue that this assumption is unrealistic.[6,7] However, RCTs require representative treatment effects rather than representative sample.[8]

The generalizability/transportability of the treatment effects depends on:[9,10]

1. Variation in the probability of enrollment in the trial

2. Heterogeneity of treatment effect or interaction

Correspondence: Oğuzhan Birdal, M.D. Department of Cardiology, Atatürk University School of Medicine, Erzurum, Turkey
e-mail: droguzhanbirdal@gmail.com

Table 1. Some basic definitions

Treatment effect	An effect attributed to a treatment in a clinical trial.
• Relative	• Odds ratio, risk ratio, hazard ratio
• Absolute	• Mean difference, risk difference, number needed to treat
External validity	Inference from trial to a target population (refer to generalizability and transportability).
Internal validity	Inference can be ascribed to differences in treatment and not confounding or baseline imbalances.
Replicability	Given a population, hypothesis, experimental design, and analysis plan, you get consistent estimates when you recollect data and redo the analysis.
Reproducibility	Given a population, hypothesis, experimental design, experimenter, data, analysis plan, and code, you get the same parameter estimates in a new analysis.
Generalizability	Inference from the trial to a target population that includes individuals who are part of the trial-eligible population.
Transportability	Inference from the trial to a target population that includes individuals who are not part of the trial-eligible population.
Interaction	The situation in which a treatment contrast is dependent on another factor.

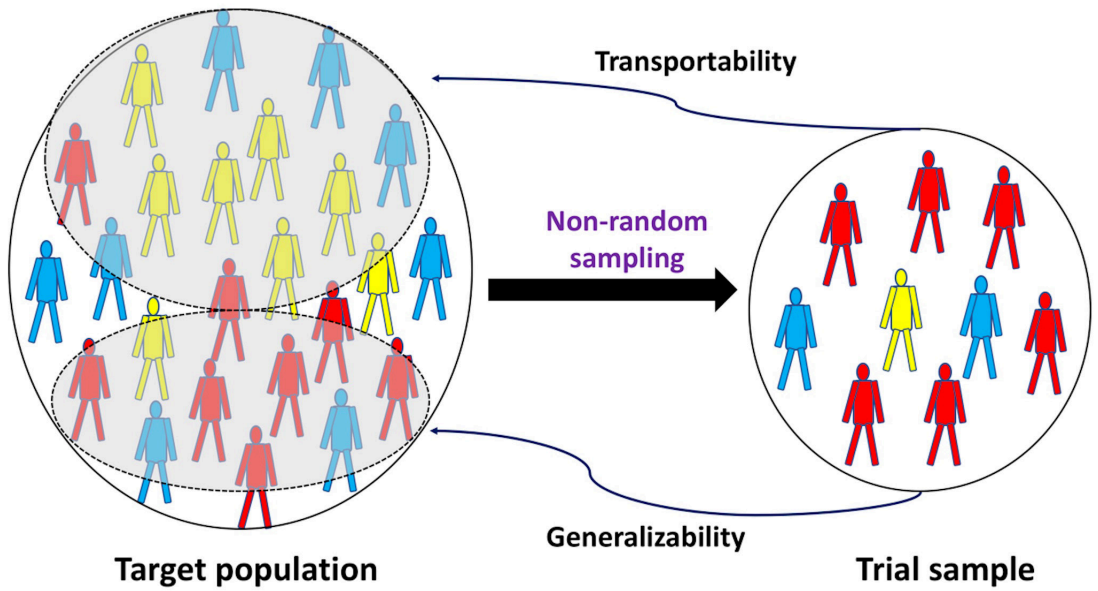

Figure 1. Transportability and generalizability of clinical trial.

Transportability of treatment effects depends on the nature of interactions. In the absence of interaction between treatment and patients' characteristics, the estimated treatment effect will apply to the target population or a much different population. However, when interaction exists, the generalization/transportability of trial results depend on the similar and sufficient distribution of interacting factors in both the RCT and the target population and allow them to be modeled to estimate treatment effects.[8]

Another important criticism against RCT is that it has strict eligibility criteria.[2] This results in many groups of patients being underrepresented (partial overlap) or not represented (no overlap) in the trial. Blacks, the elderly, or special groups (those with chronic kidney disease, cardiogenic shock, etc) can be shown as examples. Although there is no overlap or partial overlap, trial results can be generalized/transportable in the absence of interaction and in the small/simple interaction (in the latter, similar

and adequate representation and distribution of the interacting factor is taken into consideration).[7,8] Some researchers also have criticized that running an RCT under a specific protocol (patients included in RCT are followed more frequently, drug adherence is higher, drug adverse effects are monitored and detected more easily, etc) might affect the trial generalization/transportability. Using pragmatic trial designs will improve generalization/transportability of trial results.[11] Therefore, the variation in the probability of enrollment and the nature of interaction are key factors for the generalization/transportability of the trial results.

Most investigators have claimed that observational studies have weak internal validity and strong external validity. However, using observational studies and making inferences for treatment effectiveness (with multivariable regression, propensity-based methods, or instrumental variables) has a close relation to the amount of measurement accuracy of both measured and unmeasured confounders. In cases where RCT cannot be performed owing to ethical, feasibility, or time/cost issues, high-quality observational studies can be used to assess treatment effectiveness and safety.[12,13] However, observational studies often supply information on issues in clinical practice, habits, pathophysiological mechanisms, and so forth, which cannot be obtained with an RCT.

In conclusion, the generalization/transportability of the treatment effect from an RCT is much better than that from observational studies. The nature of interaction between treatment and patients' characteristics is the key to understand generalization/transportability both in RCTs and observational studies.

Conflict of interest: None.

REFERENCES

1. Hoogland J, IntHout J, Belias M, Rovers MM, Riley RD, Harrell Jr FE, et al. A tutorial on individualized treatment effect prediction from randomized trials with a binary endpoint. Stat Med 2021;40:5961-81. [Crossref]
2. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". Lancet 2005;365:82-93. [Crossref]
3. Rothwell PM. Factors that can affect the external validity of randomised controlled trials. PLoS Clin Trials 2006;1:e9. [Crossref]
4. Altın C, Topaloğlu C, Çetin N, Dalgıç O, Yavuz V, Alioğlu E, et al. Dabigatran's stroke prevention in real life; a sample of population from Turkey: D-SPIRIT registry. Turk Kardiyol Dern Ars 2021;49:630-40.
5. Connolly SJ, Ezekowitz MD, Yusuf S, Eikelboom J, Oldgren J, Parekh A, et al. Dabigatran versus warfarin in patients with atrial fibrillation. N Engl J Med 2009;361:1139-51. [Crossref]
6. Rothman KJ, Gallacher JE, Hatch EE. Why representativeness should be avoided. Int J Epidemiol 2013;42:1012-4. [Crossref]
7. Bradburn MJ, Lee EC, White DA, Hind D, Waugh NR, Cooke DD, et al. Treatment effects may remain the same even when trial participants differed from the target population. J Clin Epidemiol 2020;124:126-38. [Crossref]
8. Harrell F. Implications of interactions in treatment comparisons. Statistical Thinking. 2020. Available at: https://www.fharrell.com/post/ia/.
9. Oude Rengerink K, Kalkman S, Collier S, Ciaglia A, Worsley SD, Lightbourne A, et al. Series: Pragmatic trials and real world evidence: Paper 3. Patient selection challenges and consequences. J Clin Epidemiol 2017;89:173-80. [Crossref]
10. Degtiar I, Rose S. A review of generalizability and transportability. 2021. arXiv:2102.11904.
11. Roche N, Reddel H, Martin R, Brusselle G, Papi A, Thomas M, et al. Quality standards for real-world research. Focus on observational database studies of comparative effectiveness. Ann Am Thorac Soc 2014;11:S99-104. [Crossref]
12. Hernan MA. Methods of public health research - strengthening causal inference from observational data. N Engl J Med 2021;385:1345-8. [Crossref]
13. Lederer DJ, Bell SC, Branson RD, Chalmers JD, Marshall R, Maslove DM, et al. Control of confounding and reporting of results in causal inference studies. guidance for authors from editors of respiratory, sleep, and critical care journals. Ann Am Thorac Soc 2019;16:22-8. [Crossref]