ABC of Methodology

This is a Section of *Epidemiologia e Psichiatria Sociale*, that regularly covers methodological aspects related to the design, conduct, reporting and interpretation of clinical and epidemiological studies. We hope that these articles will help develop a more critical attitude towards research findings published in the international literature and, additionally, will help promote the implementation of original research projects with higher standards in terms of design, conduct and reporting.

Corrado Barbui, *Section Editor* and Michele Tansella, *Editor* EPS

Why internal and external validity of experimental studies are relevant for clinical practice?

ANDREA CIPRIANI, MARIANNA PURGATO, CORRADO BARBUI

Department of Medicine and Public Health, Section of Psychiatry and Clinical Psychology, University of Verona, Verona (Italy)

KEY WORDS: validity, generalisability, randomised controlled trial, statistics, statistical error.

In randomised controlled trials (RCTs) there are two types of validity: *internal validity* and *external validity*. Internal validity refers to the extent to which the observed difference between groups can be correctly attributed to the intervention under investigation. In other words, it is the extent to which the design and conduct of the trial eliminate error. Internal validity might be threatened by two types of errors: systematic error (also called *bias*) and chance error (also called *random error or statistical error*) (Keirse & Hanssens, 2000). Systematic error, or bias, may be the consequence of erroneous ways of collecting, analysing and interpreting data. This may produce differences between treatments that are not real, with an overestimation or an underestimation of the true beneficial or harmful effect of an intervention (Juni *et al*., 2001). In RCTs there are four types of bias: *selection bias* (when the groups differ in baseline characteristics because of the way participants are selected), *performance bias* (when the care provided to the trial participants differs systematically between the experimental and control group), *detection bias* (when there are systematic differences in outcome assessment), and attrition bias (when the loss of participants from the study systematically differs between the experimental and control group). By contrast, chance error, or *statistical error*, is due to outcome variability that may arise by chance alone. Studies with small sample sizes are more likely to incur in this type of error than studies with large sample sizes. Thus, the risk of random error may be minimised by recruiting sufficiently large samples of patients.

External validity refers to the extent to which trial results can be generalised to "real world" patient populations (Rothwell, 2005). The external validity of a study depends on how the study was designed and conducted. Usually, the following aspects are considered very relevant to make a judgements on the external validity of a study:

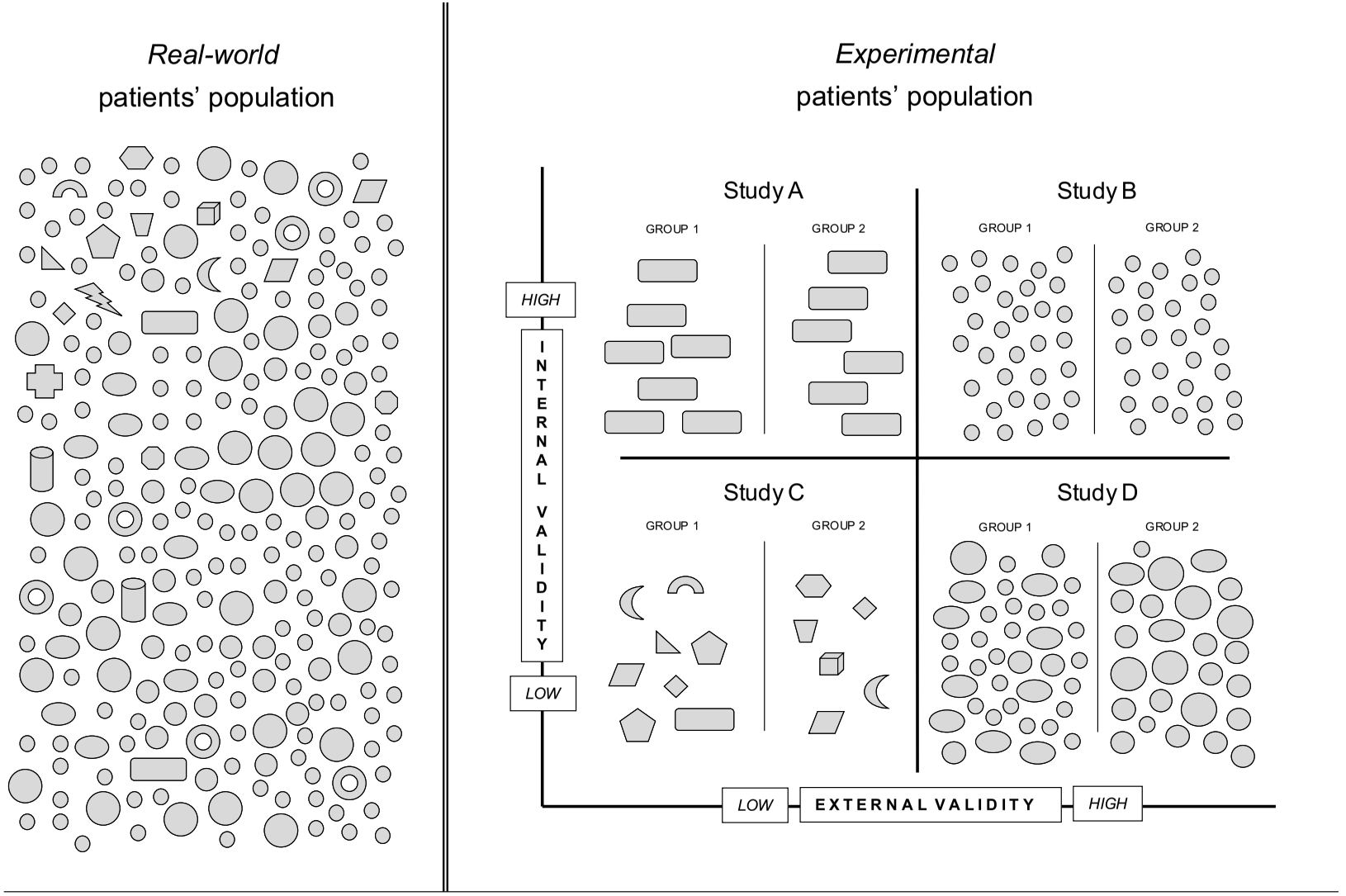
i) the setting where patients are enrolled,
ii) the inclusion/exclusion criteria adopted to select the study participants,
iii) the types of interventions used and how these are administered,
iv) the outcomes employed and how these are assessed, and
v) the duration of follow-up.

Many useful checklists have been published and can be used when critically appraising the validity of trials (Altman *et al*., 2001). Internal validity is a prerequisite for external validity. If the results of a trial are not internally valid, then the question of its external validity is irrelevant. However, a study with good internal validity

Address for correspondence: Dr. A. Cipriani, Department of Medicine and Public Health, Section of Psychiatry and Clinical Psychology, University of Verona, Piazzale L.A. Scuro 10, 37134 Verona (Italy).
Fax: +39-045-8027498
E-mail: andrea.cipriani@univr.it

https://doi.org/10.1017/S1121189X00000968 Published online by Cambridge University Press

https://doi.org/10.1017/S1121189X00000968 Published online by Cambridge University Press



Legend: Circles: patients with unipolar major depression with typical features (different sizes stand for different baseline characteristics, such as age, sex, severity of illness). Rounded rectangles:patients with unipolar major depression with atypical features. All other shapes: other types of major depression (for instance, bipolar depression or psychotic depression)

Figure 1. – *Internal and external validity of rendomised controlled trials (see text for details).*

may be clinically limited by poor external validity. If the results of a RCT are not easily applicable to real-world patients because the experimental conditions are very different from the ordinary clinical conditions (for the aspects mentioned above), then these results – even though methodologically correct – are not relevant for the target patient population (Barbui *et al*., 2007).

The representativeness of participants is particularly problematic. In clinical practice, patient populations are usually very heterogeneous, even among groups of patients with the same diagnosis. Proper randomization and reliable inclusion and exclusion criteria may help create two or more groups that do not differ. However, such groups may be different from others with the same diagnosis, for example because subject selection in many studies is restricted to patients with one typical features of the disease or to those who happen to be available or willing to participate in the trial. To exemplify this, we may want to compare the population included in a RCT with the "real world" population (Figure 1). In the Figure (left), the "real world" patient population includes those with major depression. Often, major depression have typical features, but less frequent types of depression may additionally be represented, for example atypical depression, psychotic depression, bipolar depression (Donoghue & Hylan, 2001). In the same Figure (right), 4 RCTs are described. Study A recruited an homogeneous small sample of depressed patients, poorly represented in the real world (high internal validity – low external validity); study B enrolled a homogeneous large sample of patients, highly represented in the real world (high internal validity – high external validity); study C included a small sample of patients with many different types of depression, poorly represented in the "real world" setting (low internal validity – low external validity); study D randomised a mixed sample of patients with types of depression representative of the real world population, but the two groups differed in baseline characteristics (low internal validity – high external validity). Results from RCTs are average results. In clinical practice, patients need individually tailored treatment. Statistical computations of data from RCTs give clinicians an idea of the probability that the same results would be obtained in other groups of patients with similar clinical characteristics. RCTs are the best way to know which treatment is effective but not necessarily which patients should receive that treatment.

REFERENCES

Altman D.G., Schulz K.F., Moher D., Egger M., Davidoff F., Elbourne D., Gøtzsche P.C., Lang T. & CONSORT GROUP (Consolidated Standards of Reporting Trials) (2001). The revised CONSORT statement for reporting randomised controlled trials: explanation and elaboration. *Annals of Internal Medicine* 134, 663-694.

Barbui C., Veronese A. & Cipriani A. (2007). Explanatory and pragmatic trials. *Epidemiologia e Psichiatria Sociale* 16, 124-125.

Donoghue J. & Hylan T.R. (2001). Antidepressant use in clinical practice: efficacy versus effectiveness. *British Journal of Psychiatry* 179, s9-s17.

Juni P., Altman D.G. & Egger M. (2001). Systematic reviews in health care: assessing the quality of controlled clinical trials. *British Medical Journal* 323, 42-46.

Keirse M.J. & Hanssens M. (2000). Control of error in randomized clinical trials. *European Journal of Obstetrics & Gynaecology Reproductive Biology* 92, 67-74.

Rothwell P.M. (2005). External validity of randomised controlled trials: to whom do the results apply? *Lancet* 365, 82-93.

https://doi.org/10.1017/S1121189X00000968 Published online by Cambridge University Press