This Section of *Epidemiology and Psychiatric Sciences* regularly appears in each issue of the Journal to cover methodological aspects related to the design, conduct, reporting and interpretation of clinical and epidemiological studies. The aim of these Editorials is to help developing a more critical attitude towards research findings published in international literature, promoting original research projects with higher methodological standards, and implementing the most relevant results of research in every-day clinical practice.

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

Dichotomizing rating scale scores in psychiatry: a bad idea?

M. Purgato* and C. Barbui

Section of Psychiatry, Department of Public Health and Community Medicine, University of Verona, Verona, Italy

In psychiatry, the use of rating scales as measures of outcome in clinical trials allows us to generate continuous outcome data, where each individual's outcome is measured in numbers. Continuous outcomes can be divided into two categories, such as improved and not improved, or may be kept continuous. This article briefly presents the main advantages and disadvantages of these two approaches, which are commonly employed in the analyses of rating scale scores in clinical trials and systematic reviews.

In psychiatry, the need to measure the impact of treatments on patient outcomes has led to a gradual increase in the variety of instruments available and in their use as measures of outcome in clinical trials. The use of these instruments, in the form of questionnaires or rating scales, allows us to generate continuous outcome data, where each individual's outcome is measured in numbers. Continuous data are referred to data that can take any value in a specified range, for example weight, rating scales scores, area and volume. This means that any number may be measured and reported to arbitrarily many decimals.

In terms of data management and analysis, continuous outcomes may be categorized (two categories, such as improved and not improved) or kept continuous. The aim of this ABC of Methodology is to briefly discuss the pros and cons of these two approaches,

which are commonly employed in the analyses of rating scale scores in clinical trials and systematic reviews.

Clinically, re-expressing continuous data as dichotomous can facilitate understanding and applicability of results, as it allows doctors to express in terms of the proportion of patients and not in terms of means and standard deviations (Table 1). In clinical trials and meta-analyses of trial data, categorization of continuous outcome measures allows us to express differences between competing treatments in terms of risk difference, relative risk or odds ratio, which are commonly employed and relatively easy to understand measures of treatment effect. However, dichotomizing leads to several problems (Table 1) (Altman & Royston, 2006). A first issue is that it may seriously underestimate the extent of variation in outcome between groups, losing information and statistical power. This may increase the risk of type I error, which means failing to detect a difference that is real, a major drawback in clinical trials and meta-analyses. A second issue is that the definition of the cut-point may not be a straightforward task, and may be rather arbitrary and not based on any solid clinical reasoning.

* Address for correspondence: Dr Marianna Purgato, Section of Psychiatry, Department of Public Health and Community Medicine, University of Verona, Piazzale L.A. Scuro, 10-37134 Verona, Italy.

(Email: marianna.purgato@univr.it)

Table 1. *The pros and cons of re-expressing continuous outcome measures as dichotomous*

Categorization of continuous variables			
Yes		No	
Pros	Cons	Pros	Cons
In clinical practice, it is helpful to label individuals as having or not having an outcome. It allows physicians to express in terms of the proportion of patients. Differences between competing treatments can be presented as RD, RR, OR and NNT.	Much information is lost, statistical power is decreased. Increases the probability of detecting false positive results – type I error. Cut-point difficult to establish, very arbitrary. Individuals close to but on opposite sides of the cut-point are categorized as very different rather than very similar. Small differences in means may be artificially translated into large differences in proportions.	All information is used. Takes into account the possibility that outcome measures might have a heterogeneous distribution. Small mean differences may have a significant impact on many patients.	In clinical practice, it is counterintuitive to express in terms of means and standard deviations. The clinical impact of small differences is unclear. In meta-analysis of clinical trial data, lumping together mean scores from different rating scales poses problems in terms of statistical assumptions and clinical interpretation.

RD, risk difference; RR, relative risk; OR, odds ratio; NNT, number needed to treat.

Consequently, it may happen that individuals close to, but on opposite sides of the cut point are considered as very different rather than very similar, which is clinically counterintuitive (Table 1). A third issue is the possibility that re-expressing continuous data as dichotomous may artificially produce large differences in proportions. Moncrieff & Kirsch (2005), who hypothesized a situation of one point difference in mean change of scores on the Hamilton rating scale between drug and placebo, showed that defining response as a minimum 12-point improvement on the Hamilton rating scale for depression, if improvement is normally distributed and the criterion for response is close to the mean improvement rate, a response rate of 50% in the drug condition and 32% in the placebo condition can be obtained.

In contrast, the main advantage of keeping data continuous is that all available information is used (Table 1). This is of paramount importance, as even a small difference between two means may have a significant impact on many patients. Guyatt *et al.* hypothesized a situation of a randomized clinical trial showing a mean difference of 0.25 in a questionnaire in which the minimal important difference is 0.5 (Guyatt *et al.* 1998). It may be erroneously concluded that the difference is clinically not relevant, but this interpretation would be based on the assumption that every patient treated scored 0.25 better than they would have scored had they received the control treatment. This would ignore the possibility that treatment might have a heterogeneous effect and, depending on the true distribution of results, the appropriate interpretation might be different. Therefore, keeping data continuous may help identify heterogeneity in treatment effect (Table 1). At the same time, however, this can lead to several problems (Table 1). The first concern is that in clinical practice it is counterintuitive to express in terms of means and standard deviations, as doctors treat individual patients. The clinical meaning of differences in means may be rather difficult to extrapolate, as mean differences are not easily translated into proportions of patients who may benefit. In meta-analyses of trial data, additionally, the need for lumping together means and standard deviations from different rating scales has led to the use of standardized mean differences, which are an artefact as these standardized measures apply to a theoretical reference rating scale that does not exist in real life.

We argue that critical appraisal of findings from randomized clinical trials and systematic reviews should consider how continuous outcome data from rating scales have been manipulated and analyzed. Physicians should be encouraged to interpret study findings taking into consideration all possible

implications of re-expressing continuous data as dichotomous versus keeping them continuous. Physicians should also be aware that it is possible to design clinical trials (Lieberman *et al.* 2005) and systematic reviews (Barbui *et al.* 2008) that, instead of relying on rating scale scores as primary outcome measures, employ pragmatic outcomes (Barbui *et al.* 2007), such as suicide attempts, treatment switching, hospitalization, school failure or truancy, job loss, or even dropping out of the trial itself. These outcomes have the added value of being very close to real life without requiring any form of manipulation.

References

Altman DG, Royston P (2006). The cost of dichotomising continuous variables. *British Medical Journal* **332**, 1080.

Barbui C, Veronese A, Cipriani A (2007). Explanatory and pragmatic trials. *Epidemiology and Psychiatric Sciences* **16**, 124–125.

Barbui C, Furukawa TA, Cipriani A (2008). Effectiveness of paroxetine in the treatment of acute major depression in adults: a systematic re-examination of published and unpublished data from randomized trials. *Canadian Medical Association Journal* **178**, 296–305.

Guyatt GH, Juniper EF, Walter SD, Griffith LE, Goldstein RS (1998). Interpreting treatment effects in randomised trials. *British Medical Journal* **316**, 690–693.

Lieberman JA, Stroup TS, McEvoy JP, Swartz MS, Rosenheck RA, Perkins DO, Keefe RS, Davis SM, Davis CE, Lebowitz BD, Severe J, Hsiao JK (2005). Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *New England Journal of Medicine* **353**, 1209–1223.

Moncrieff J, Kirsch I (2005). Efficacy of antidepressants in adults. *British Medical Journal* **331**, 155–157.