

Balanced or imbalanced samples?

MEDICINE AND NUMBERS

STIAN LYDERSEN

E-mail: stian.lydersen@ntnu.no

Stian Lydersen, PhD and professor of medical statistics at the Regional Centre for Child and Youth Mental Health and Child Welfare (RKBU), Department of Mental Health, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU).

The author has completed the ICMJE form and reports no conflicts of interest.

When comparing two groups, the groups are usually planned to be equally large. But moderate imbalance need not cause a notable reduction in statistical power. And in some settings, imbalance can give increased statistical power.

In randomised controlled trials, the groups are usually planned to be equally large. But some randomisation procedures, particularly in multicentre trials, may result in one group being somewhat larger than the other, even with 1:1 randomisation. However, this does not present a problem, since statistical power is not notably reduced with moderate imbalance (1, p, 46). In other studies, such as case-control studies, the number of subjects in one group may be limited. The power may then be increased to some extent by including more subjects in the other group.

When is moderate imbalance not problematic?

Consider a randomised controlled trial planned with two groups of 100 patients, that is, a total of 200 patients. With a normally distributed outcome variable and an effect size equal to 0.4 standard deviations, the statistical power will be 80.4 % at significance level 5 %. If, instead, 110 and 90 patients are included in the respective groups, which is a large imbalance by chance with 1:1 randomisation, the power will be 80.0 % under the same

assumptions. Such an imbalance therefore has a negligible effect on power. Even a 2:1 randomisation with 133 and 67 patients in the two groups would result in a modest reduction in power to 75.5 %. Only a large imbalance would give a substantial reduction in power. For example, group sizes of 150 and 50 would result in a power of only 64.8 % (Figure 1).

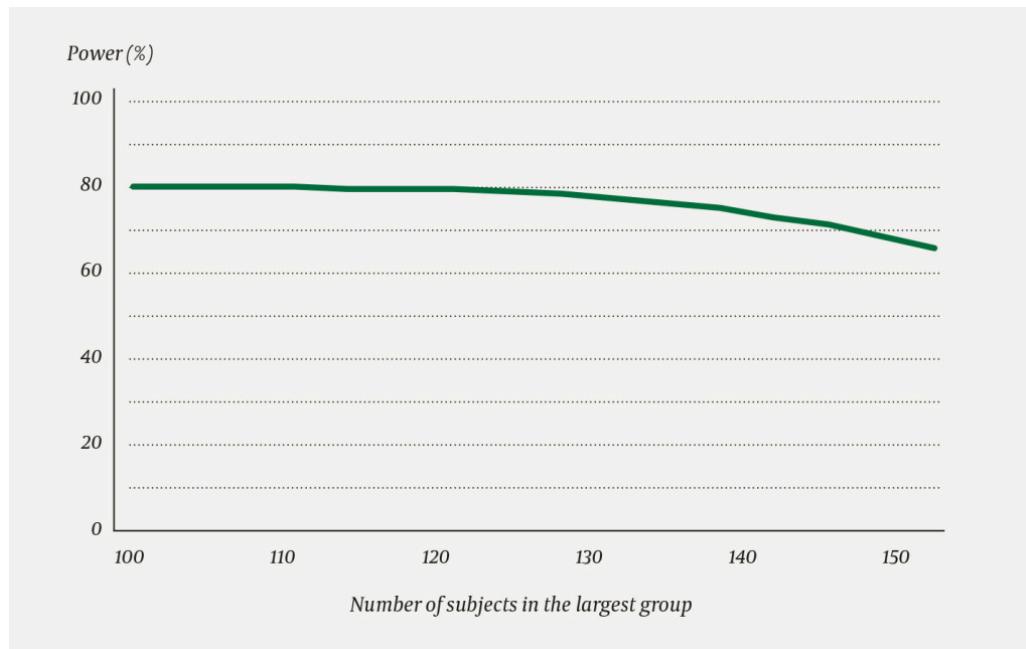


Figure 1 Statistical power for comparison of two groups using the t-test. The total number of subjects in the two groups is 200. Computed for effect size 0.4 standard deviations and significance level 5 %.

When is imbalance beneficial?

In some situations, the number of patients in the treatment group may be limited, for example due to high treatment costs, while including more patients in the control group may be less costly or easier. Statistical power can be increased to some extent in this case by increasing the number of patients in the control group. Let us assume only 70 patients can be included in the treatment group. With 70 patients in the control group, the power would be only 65.2 % under the same assumptions as above. Usually, the aim is for a statistical power of at least 80 % for a trial. Increasing the control group to 140 gives a power of 77.6 %, and a control group of 210 gives a power of 82.3 %.

A similar issue may also be relevant in a case-control study of a rare disease, where few patient cases are available. The control group consists of persons without the disease. The number of those subjected to a specific exposure is recorded for each group in order to investigate whether there is an association between the exposure and the disease. Let us assume that only 100 cases are available, and that we expect that 10 % of these and 2 % of the controls have been exposed. If we include 100 controls in the study, the statistical power will be only 66.6 % at a significance level of 5 %. Increasing the number of controls to 200 or 400 gives a power of 82.4 % or 89.8 %, which is usually considered sufficient to undertake a study. This is illustrated in Figure 2. A rule of thumb

says that little is gained by increasing the largest group to more than four to five times the smallest group. (1, p. 124). In Figure 2, the graph levels out around 400 to 500 in the largest group.

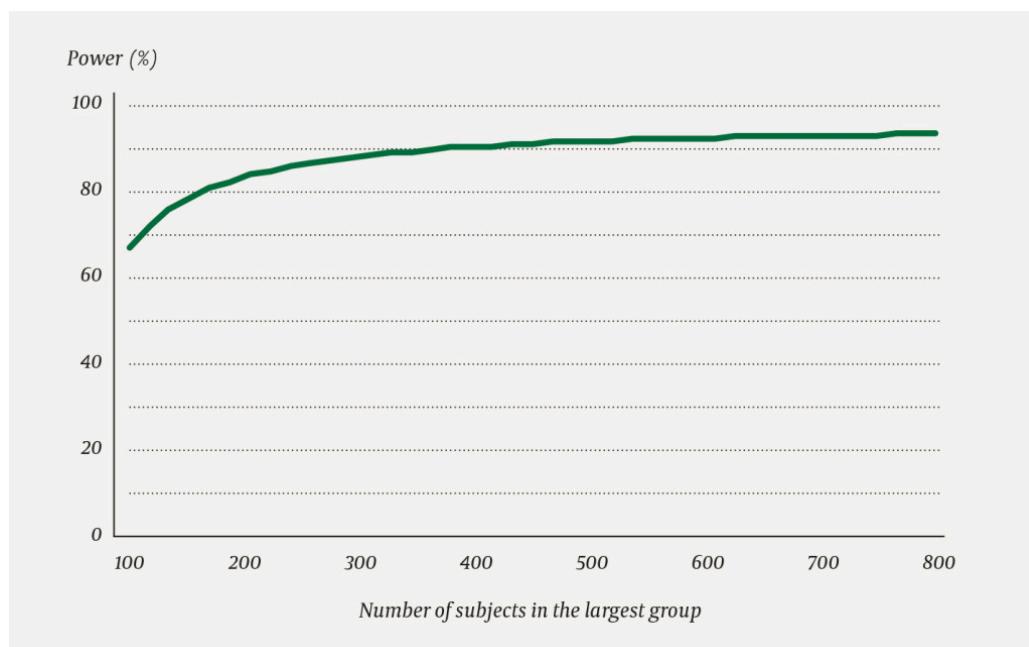


Figure 2 Statistical power for comparing two groups using Pearson's chi-squared test at significance level 5 %. In this example, the smallest group size is 100, and the probability in the smallest and largest group is 10 % and 2 %, respectively.

Conclusion

We have examined two settings: When the total number of subjects in a trial is given, the power will not be notably reduced by some imbalance between the groups. When the number in one group is limited, the power can be somewhat increased by increasing the number of subjects in the other group.

LITERATURE

1. van Belle G. 2008. Statistical Rules of Thumb, 2 ed. Hoboken, New Jersey, Wiley.
