Attrition analysis

STIAN LYDERSEN

stian.lydersen@ntnu.no
Stian Lydersen, PhD, professor of medical statistics at the Regional Centre for Child and Youth Mental Health and Child Welfare, Department of Mental Health, Norwegian University of Science and Technology.
The author has completed the ICMJE form and declares no conflicts of interest.

Attrition will occur in most clinical studies. If those who drop out differ substantially from those who participate, the generalisability of the study's results may suffer.

Attrition can occur because not all of the invited participants choose to take part in the study, or because some of the included participants fail to attend the follow-up.

Descriptive statistics

In an attrition analysis, participants who have dropped out of the study are compared to those who were included in or completed it. For example, the Health Survey in the Department of Children and Youth, Division of Mental Health Care, St Olav's Hospital (St Olav CAP Survey) is a longitudinal cohort study in which all those who had been outpatients and inpatients at the clinic in the period 2009–11 were invited to take part. Those who were included were also invited to follow-up studies three and nine years later. A total of 1 743 adolescents were eligible for inclusion, and 717 of them consented and were included in the study. It turned out that these 717 were slightly older, with a mean age of 15.66 years (standard deviation 1.65), than those who were not included, who had a mean age of 15.39 years (standard deviation 1.95). The proportion of girls was also higher, with 54.8 % compared to 49.6 % (1).

P-values are of little relevance

Mangerud et al. also reported that these differences were statistically significant, with *p*-values of 0.0015 and 0.032 respectively ([1]). However, it is difficult to see how *p*-values in attrition analyses can be relevant, although unfortunately they are often requested and reported. It will rarely be of importance whether the differences are statistically significant. A difference of 0.27 years was hardly of any practical importance in the study in question, but it became highly significant statistically because of the large number of persons in both groups. Similarly, large differences between the groups can be observed in smaller studies, but without these being statistically significant. The relevant issue is whether this difference can be regarded to be of practical importance. Later publications from the Hel-BUP study chose to only report descriptive statistics in the attrition analyses, such as in Table 1, based on Gårdvik et al. ([2]). This is in line with the recommendations in the 'Vancouver guidelines', which place less emphasis on *p*-values than previously ([3]). In an attrition analysis, what is most important is to report descriptive statistics for participants and non-participants separately. In addition to numbers, this will normally include means and standard deviations for continuous data, and counts and proportions for categorical data, as in Table 1.

Table 1

Attrition analysis from inclusion to three-year follow-up in the Health Survey in the Department of Child and Adolescent Psychiatry, St Olav's Hospital. From Table S1 in Gårdvik et al. (2). SD = standard deviation.

	Participants at inclusion (*n* = 717)	Participants after three years (*n* = 570)	Non-participants after three years (*n* = 147)
Age in years at inclusion, average (SD)	15.7 (1.7)	15.7 (1.7)	15.5 (1.6)
Girls, proportion (%)	393 (54.8)	324 (56.8)	69 (46.9)

Handling missing data

The CAP Survey had been granted permission by the Regional Committee for Medical and Health Research Ethics to record age, sex and reason for referral for those who were not included in the study. This enabled comparisons between those who were included and those who were not. When such data are available, bias can be reduced by, for example, weighting the subsequent analyses. However, such a permission tends to be granted as the exception rather than the rule.

A common reason for attrition is that participants fail to attend the follow-up. In such cases, it could be relevant to include in the analysis available data for participants who have dropped out. Data are rarely missing completely at random (4). By including participants with partially missing data we can avoid or reduce bias caused by differences between participants and non-participants. Moreover, statistical power will be somewhat higher, leading to increased precision in the results. There are alternative ways to do this. In a follow-up study, a mixed model could be suitable (5). In other cases, multiple imputation of missing data could be an option (6). In some cases, the 'full information maximum likelihood' method can be used, but this is extremely computationally intensive and not always feasible (4).

Summary

Attrition analysis should be based on descriptive statistics. In certain cases, participants with some missing data can be included in the analysis using the data that are available.

REFERENCES

1. Mangerud WL, Bjerkeset O, Lydersen S et al. Chronic pain and pain-related disability across psychiatric disorders in a clinical adolescent sample. BMC Psychiatry 2013; 13: 272. [PubMed][CrossRef]

2. Gårdvik KS, Rygg M, Torgersen T et al. Psychiatric morbidity, somatic comorbidity and substance use in an adolescent psychiatric population at 3-year follow-up. Eur Child Adolesc Psychiatry 2021; 30: 1095–112. [PubMed][CrossRef]

3. Lydersen S. Færre p-verdier i «Vancouver-retningslinjene». https://tidsskriftet.no/2020/01/kommentar/faerre-p-verdier-i-vancouver-retningslinjene#comment-1990 Accessed 19.9.2022.

4. Lydersen S. Manglende data – sjelden helt tilfeldig. Tidsskr Nor Legeforen 2019; 139: 269. [CrossRef]

5. Lydersen S. Analyse av longitudinelle data. Tidsskr Nor Legeforen 2022; 142: 416. [CrossRef]

6. Lydersen S. Multippel imputering av manglende data. Tidsskr Nor Legeforen 2022; 142: 151. [CrossRef]