

Studies of Attrition in Rheumatological Databases



Whereas clinical trials are the method of choice to answer many medical questions, especially those related to the comparison of treatments, patient registries have an important role in the study of chronic diseases. However, the resources that are available for patient followup are inevitably more limited for a registry database than for a clinical trial. The potential impact of losses to followup must therefore be considered when making use of registry data.

In mortality studies, for example, a standard procedure is to define censoring times, the time beyond which death is known to occur for patients not observed to die, as the date that patients were last seen alive. This supposes that losses to followup are independent of the risk of death, an assumption that is usually not completely true. Thus this standard assumption may bias any study of mortality. The ability to deal with such bias depends on knowledge about patterns of patient attrition from a database.

Studies of attrition in epidemiological cohorts are common but, as Krishnan, *et al* point out in this issue of *The Journal*¹, attrition patterns in clinical databases are less well understood. However, some studies have been done, one from a psoriatic arthritis clinic², another from a lupus clinic³, and most recently Krishnan, *et al*¹ have provided a valuable look at the ARAMIS (Arthritis, Rheumatism and Aging Medical Information System) rheumatoid arthritis (RA) database. The study of attrition in this database is particularly important because it represents such a large number of patients in different settings.

While it is important to quantify the amount of attrition in a registry, a key benefit of specific studies of attrition is the identification of factors related to the probability of being lost to followup. In the RA study¹, key factors identified were younger age, lower levels of education, and non-Caucasian race. Such knowledge can help to establish registry procedures that help minimize bias. Oversampling is one approach, but this option may not always be available, for example in a clinic setting. Simply the awareness

that some patients are at a particular risk of being lost to followup, however, may help to reduce the risk.

Many correlated factors will be related to attrition. Thus it is important, as recognized by Krishnan, *et al*¹, to make use of multivariate analyses to identify the most important contributors to attrition. In a metaanalysis of attrition in longitudinal studies in the elderly⁴, it was noted that many factors were related to attrition in univariate analyses, but a consistent pattern across studies only emerged in multivariate analyses.

Knowledge of risk factors for attrition may be sufficient to allow valid analyses of longitudinal data. Technically, this is often true if a missing at random (MAR) assumption can be maintained. This assumption means that conditional on the observed data, including known risk factors, the probability that data are missing does not depend on unobserved data, in particular the response variable of interest for those lost to followup. More generally, however, models that handle the difficult situation of dependent or non-ignorable dropouts involve assumptions that can only be checked by obtaining supplementary information on dropouts. Surprisingly, therefore, patient tracing⁵ is less commonly attempted in attrition studies. If any potential for special one-time efforts to trace patients exists, then it may add considerably to the understanding of the relationship between attrition and outcomes. For example, in a tracing study among lupus patients³, it was shown that recent clinic attendance was not demonstrably related to mortality, thus justifying the use of the registry data for mortality studies.

While the value of studies of attrition is considerable, they present particular problems for data analysis. In a longitudinal database, even the definition of attrition is problematic. Krishnan, *et al*¹ focus on the completion of a Health Assessment Questionnaire (HAQ) at data collection cycles and define attrition to be “non-completion of the last HAQ mailed to the patient at the cutoff date, the 38th mailing cycle in 1999.” They acknowledge, however, that some patients fill in questionnaires erratically over time, so

See Attrition bias in rheumatoid arthritis databanks: A case study of 6346 patients in 11 databanks . . . , page 1320

that patients are “not considered dropouts so long as they were known to be alive at the cutoff date and completed the last questionnaire at that time.” This pragmatic and simple definition has a certain attraction. However, it may present difficulties in analysis. Krishnan, *et al*¹ use time-to-event methodology and define the time to being lost to followup as the date of completion of the last HAQ. Problems of interpretation then arise. For example, from a plot of the percentage of patients retained by collection cycle, about 75% of patients are estimated to be retained after 5 cycles and 50% after 20 cycles. However, the value of 75% refers to the probability of not completing a HAQ at or before the 5th cycle AND not subsequently completing a HAQ for 33 cycles, whereas the value of 50% refers to the probability of not completing a HAQ at or before the 20th cycle AND not subsequently completing a HAQ for 18 cycles.

The additional condition of continuing noncompletion of HAQ over periods of time that vary from patient to patient also creates technical difficulties in the analysis. It is not immediately clear that the usual statistical methods for time-to-event data can be used in this situation since it is not possible to determine, in principle at least, whether an event has occurred at the supposed time of occurrence.

This potential problem with such time-to-event analyses is symptomatic of many longitudinal studies of attrition, the analysis of which must involve consideration of both the outcome process of primary interest and some sort of ascertainment process⁶. In a lupus tracing study, a rather careful development⁵ was required to justify the use of apparently straightforward time-to-event methodology. In addition, while the validity of the test of a null hypothesis of no relationship between lost to followup status and mortality through use of an apparent relative-risk parameter was established, the implied family of models could not be regarded as interpretable representations of non-null relationships. For the study reported in Krishnan, *et al*¹, it would perhaps be possible to justify the combination of (uncorrelated but not independent) cross-sectional analyses by cycle to test null hypotheses, but parameter estimation may remain difficult to interpret. Analysis by cycle might also be informative if the risk factors for attrition vary with cycle, a possibility discussed by Deeg⁷.

Usually, consideration of technical statistical details is not appropriately published in a medical journal, but should underlie any presentation of results. Also, as in other settings, it is sometimes sensible to publish a less than ideal, but valid, analysis to ensure trial results are available in a timely fashion. However, in such situations, it is valuable to carefully specify any simplifying assumptions and to

provide sensitivity analyses relevant to these assumptions. Such analyses may have helped to firmly establish the conclusions drawn by Krishnan, *et al*¹ from the valuable ARAMIS data, even if questions remained about the primary analysis.

For rheumatological diseases, there are many questions that can only be addressed through the use of clinical databases. The increasing recognition of this is leading to the establishment of such databases and this should be encouraged and appropriately funded by research organizations. It is important to present sensitivity analyses of the effect of attrition on inferences from these databases. Also, more formal studies of attrition are needed, and it is important to establish procedures for the appropriate analysis of attrition. However, since the key aim is to ensure that attrition does not reduce the usefulness of databases, the most important result of attrition studies might be simply to encourage, and facilitate, the best possible followup of patients that resources allow.

VERNON T. FAREWELL, PhD,
Senior Scientist, MRC Biostatistics Unit,
Institute of Public Health,
University Forvie Site,
Robinson Way,
Cambridge, UK CB2 2SR.

Address reprint requests to Dr. Farewell.

REFERENCES

1. Krishnan EK, Murtagh K, Bruce B, Cline D, Singh G, Fries JF. Attrition bias in rheumatoid arthritis databanks: A case study of 6346 patients in 11 databanks and 65,649 administrations of the Health Assessment Questionnaire. *J Rheumatol* 2004;31:1320-6.
2. Brubacher B, Gladman DD, Buskila D, Langevitz P, Farewell VT. Follow-up in psoriatic arthritis: Relationship to disease characteristics. *J Rheumatol* 1992;19:917-20.
3. Gladman DD, Koh DR, Urowitz MB, Farewell VT. Lost to follow-up study in SLE. *Lupus* 2000;9:363-7.
4. Chatfield MD, Brayne CE, Matthews FE. Systematic literature review of attrition between waves in large population based longitudinal studies in the elderly. *J Clin Epidemiol* 2004; (in press).
5. Farewell VT, Lawless JF, Gladman DD, Urowitz MB. Tracing studies and analysis of the effect of loss to follow-up on mortality estimation from patient registry data. *Appl Statist* 2003;52:445-56.
6. Hu P, Tsiatis AA. Estimating the survival distribution when ascertainment of vital status is subject to delay. *Biometrika* 1996;83:371-80.
7. Deeg DJH. Attrition in longitudinal population studies: Does it affect the generalizability of the findings? *J Clin Epidemiol* 2002;55:213-5.