# Longitudinal Studies in Rheumatology
## Some Guidance for Analysis

**Emmanuel Lesaffre, Dr.Sc.**

***Abstract***

*In a follow-up study, patients are monitored over time. Longitudinal and time-to-event studies are the two most important types of a follow-up study. In this paper, the focus is on longitudinal studies with a continuous response where patients are examined at several time points. While longitudinal studies provide a powerful tool for the evaluation of a treatment effect over time, a major problem is missing data caused, for example, by patients who drop out from the study. Many longitudinal studies in rheumatology use inappropriate statistical methodology because either they do not address correctly the correlated nature of the repeated measurements, or they treat the problem of missing data incorrectly. We will illustrate that there are interpretational and computational issues with the "classical" approaches. Further, we expand here on more appropriate statistical techniques to analyze longitudinal studies. To this end, we focus on randomized controlled trials (RCTs) and illustrate the approaches on data from a fictive randomized controlled trial in rheumatology.*

In a follow-up study, patients are monitored over time. In a time-to-event study, the interest lies in recording the time until an event occurs. When the event is mortality, one speaks of a survival study. While the patient's condition is monitored over time, (e.g. for safety reasons) in this type of follow-up study the repeated evaluations of the patient are not of primary importance; what counts is the time to the event.

On the other hand, *a longitudinal study*, also called *repeated measurements study*, is needed when the primary outcome of the study needs to be monitored repeatedly over time. In contrast to a study where the clinical examinations are done only at baseline and at the end of the study, the longitudinal follow-up allows to monitor the rate of change in the response. This study type provides also a better protection against the harmful effect of missing data, as will be seen below.

Of key importance to choose the appropriate statistical approach is the nature of the *missing data process*. When missing data occur during the conduct of the study but the patient stays in the study, the missing data are called *intermittently*. But when the patient decides to leave the study, one says that the patient *drops out* from the study, and this missing data process is, therefore, called a *dropout process*. In the next section, we distinguish between three types of missing data/dropout processes. We illustrate on a fictive study what the effect these dropout processes may have on descriptive and inferential statistics. We also show that the classical statistical approaches fail to take into account appropriately missing data.

Emmanuel Lesaffre, Dr.Sc., is Chair of the Department of Biostatistics, Erasmus MC, Rotterdam, The Netherlands.
*Correspondence:* Emmanuel Lesaffre, Dr.Sc., Department of Biostatistics, Erasmus MC, Rotterdam, The Netherlands; e.lesaffre@erasmusmc.nl.

## Why Missing Data Occur or Why Do Patients Drop Out from a Study?

There are a variety of reasons why data fail to be collected in a study, ranging from pure bad luck to a process that is intimately related to the disease of the patient or the intervention that is administered. Below we describe a classical taxonomy that was introduced by Little and Rubin.[1,2] The classification determines the choice of the appropriate statistical approach, but the terminology is somewhat confusing. It is assumed that the response of interest may be lacking, but that other measurements (appearing possibly as covariates in a regression model) are available.

*Missing-completely-at-random (MCAR):* A missing response occurs because of reasons completely unrelated to the response (i.e., by pure bad luck). This happens for instance

Lesaffre E. Longitudinal studies in rheumatology: some guidance for analysis. Bull NYU Hosp Jt Dis. 2012;70(2):65-72.