Recent development in the handling of missing data
Mike Kenward (London School of Hygiene & Tropical Medicine)

Date: 8 June, 2006
Venue: Room 1.25/1.26, Kirkland House

Missing data can be defined as those observations whose collection was intended but which, for one reason or another, were not obtained. There are two main implications of missing data for the subsequent statistical analysis. The first is largely computational: how should the consequent lack of balance been managed in the analysis? The second, which is probably the more important: what are the consequences for the efficiency and validity of the analysis? One common approach to analysis is to only use data from units with complete sets of measurements. It is clear that in data sets with many explanatory variables and potentially missing values spread among most of these, such a route can be hugely inefficient. Further it is quite possible that, in the set of completers analysed, properties of variables, and the relationships among them, may be quite different from those in the complete sample. Multiple imputation, developed nearly thirty years ago by Donald Rubin, is a very important tool for dealing with both the issues of computation and, to some extent, bias with large complex sets of incomplete data. In this talk an introduction will be given to the method. Links to other methods, advantages, disadvantages, and key remaining issues with the technique will be discussed.