

Piracicaba, SP, Brazil, 2006

Course

Title of course:

Longitudinal and Incomplete Data

Names and addresses of presenters:

GEERT VERBEKE (50%), Biostatistical Centre, K.U.Leuven,
Kapucijnenvoer 35, B-3000 Leuven, Belgium,
Email: geert.verbeke@med.kuleuven.be,
Tel: +32-16-336891, Sec: +32-16-336892, Fax: 32-16-337015,
Web: <http://www.kuleuven.be/biostat>.

GEERT MOLENBERGHS (50%), Center for Statistics, Universiteit Hasselt,
Agoralaan 1, Building D, B-3590 Diepenbeek, Belgium,
Email: geert.molenberghs@uhasselt.be,
Tel: +32-11-268238, Sec: +32-11-268202, Fax: +32-11-268299,
Web: <http://www.uhasselt.be/censtat>.

Abstract:

We first present linear mixed models for continuous hierarchical data. The focus lies on the modeler's perspective and on applications. Emphasis will be on model formulation, parameter estimation, and hypothesis testing, as well as on the distinction between the random-effects (hierarchical) model and the implied marginal model. Apart from classical model building strategies, many of which have been implemented in standard statistical software, a number of flexible extensions and additional tools for model diagnosis will be indicated. Second, models for non-Gaussian data will be discussed, with a strong emphasis on generalized estimating equations (GEE) and the generalized linear mixed model (GLMM). To usefully introduce this theme, a brief review of the classical generalized linear modeling framework will be presented. Similarities and differences with the continuous case will be discussed. The differences between marginal models, such as GEE, and random-effects models, such as the GLMM, will be explained in detail.

Third, when analysing hierarchical and longitudinal data, one is often confronted with missing observations, i.e., scheduled measurements have not been made, due to a variety of (known or unknown) reasons. It will be shown that, if no appropriate measures are taken, missing data can cause seriously jeopardize results, and interpretational difficulties are bound to occur. Methods to properly analyze incomplete data, under flexible assumptions, are presented. Key concepts of sensitivity analysis are introduced.

Throughout the course, it will be assumed that the participants are familiar with basic statistical modelling, including linear models (regression and analysis of variance), as well as generalized linear models (logistic and Poisson regression). Moreover, pre-requisite knowledge should also include general estimation and testing theory (maximum likelihood, likelihood ratio).

All developments will be illustrated with worked examples using the SAS System. These will be supplemented with practical sessions.

Presenters:

GEERT VERBEKE is Professor in Biostatistics at the Biostatistical Centre of the Katholieke Universiteit Leuven in Belgium. He received the B.S. degree in mathematics (1989) from the Katholieke Universiteit Leuven, the M.S. in biostatistics (1992) from the Limburgs Universitair Centrum, and earned a Ph.D. in biostatistics (1995) from the Katholieke Universiteit Leuven. Dr Verbeke wrote his dissertation, as well as a number of methodological articles, on various aspects of linear mixed models for longitudinal data analyses. He has held visiting positions at the Gerontology Research Center and the Johns Hopkins University (Baltimore, MD). Geert Verbeke is Past President of the Quetelet Society, International Program Chair for the International Biometric Conference in Montreal (2006), and Joint Editor of the *Journal of the Royal Statistical Society, Series A* (2005–2008).

GEERT MOLENBERGHS is Professor of Biostatistics at the Universiteit Hasselt in Belgium. He received the B.S. degree in mathematics (1988) and a Ph.D. in biostatistics (1993) from the Universiteit Antwerpen. Dr Molenberghs published methodological work on surrogate markers in clinical trials, categorical data, longitudinal data analysis, and on the analysis of non-response in clinical and epidemiological studies. He serves as Joint Editor for *Applied Statistics* (2001–2004) and is President of the International Biometric Society (2004–2005). He was elected Fellow of the American Statistical Association and received the Guy Medal in Bronze from the Royal Statistical Society. He has held visiting positions at the Harvard School of Public Health (Boston, MA).

Both presenters are editor and author of three books on the use of linear mixed models for the analysis of longitudinal data (Springer Lecture Notes 1997, Springer Series in Statistics 2000, Springer Series in Statistics 2005), and they have taught numerous short and longer courses on the topic in universities as well as industry, in Europe, North America, Latin America, and Australia.

Course materials:

- ▷ Copies of the transparencies used in the course
- ▷ Verbeke, G. and Molenberghs, G. (2000) *Linear Mixed Models for Longitudinal Data*. New York: Springer-Verlag.
- ▷ Molenberghs, G. and Verbeke, G. (2005) *Models for Repeated Discrete Data*. New York: Springer-Verlag.

Learning outcomes and instructional methods:

As a result of the course, participants should be able to perform a basic analysis for a particular longitudinal data set at hand, using linear, generalized linear, and non-linear tools for longitudinal data. Based on a selection of exploratory tools, the nature of the data, and the research questions to be answered in the analyses, they should be able to construct an appropriate statistical model, to fit the model within the SAS framework, and to interpret the obtained results. Further, participants should be aware not only of the possibilities and strengths of a particular selected approach, but also of its drawbacks in comparison to other methods.

The course will be explanatory rather than mathematically rigorous. Emphasis is on giving sufficient detail in order for participants to have a general overview of frequently used approaches, with their advantages and disadvantages, while giving reference to other sources where more detailed information is available. Also, it will be explained in detail how the different approaches can be implemented in the SAS package, and how the resulting outputs should be interpreted.