

14/08/2000

**James Rennie Bequest Travel Grant for the
International Biometrics Conference 2000**

Florence Jaffrézic, PhD student at ICAPB Edinburgh University.

The International Biometrics Conference gathered researchers from all over the world to discuss the advancements made in biological science through the development and application of new mathematical and statistical techniques. One of the main issues this year was the analysis of longitudinal data, a topic central to my PhD. In fact, the purpose of my work is to propose and compare models for the genetic evaluation of dairy cattle for milk production using test day records.

The course on longitudinal data analysis, given by two very respected researchers in the area, namely K-Y Liang and S. L. Zeger, was extremely interesting. It offered a very good overview of methodologies that have been used in this area, and presented new avenues of research in the subject.

During the conference, many presentations also dealt with the topic of missing data in longitudinal studies (R. Little), joint modelling of longitudinal and event time data (P. Diggle), or comparison of statistical models for the analysis of continuous longitudinal data (G. Verbeke), illustrating nicely the points presented during the course as well as introducing new methodologies.

Many other topics currently of interest in biometrical research were also presented, for example survival analysis, geostatistics, and genetics with quantitative trait loci detection and marker assisted selection. The very rich program with high quality presentations allowed us to widen our view on the Biometrics research and to meet people confronting similar statistical problems but working in different fields.

The conference was also an excellent opportunity for me to present the work I completed during the first year of my PhD to a very competent public, who could give interesting and relevant feedback. This enabled me to establish new contacts with researchers working on similar topics, and to discuss some issues with them.

This experience was therefore very beneficial as the Conference highlighted new avenues of research relevant to my PhD work.

Main issues in longitudinal data analysis

Three main issues were discussed during Liang and Zeger's course and the conference about longitudinal data analysis.

1. Missing data and modelling the drop-out process

Longitudinal data models are often said to be able to accommodate missing data. This means in fact that they are able to accommodate unbalanced data. However, a common phenomenon in longitudinal data is the 'drop-out' process, meaning that some individuals terminate prematurely (for example, some cows can be dried off earlier than others). This kind of missing data, when the drop-out process is not independent from the process of interest (for example milk production), is not taken into account by classical longitudinal models like random regression, or character process models. Two main methodologies have been proposed to model the informative drop-out process:

- **Selection model** (Diggle and Kenward, 1994): the drop-out process is modelled using, for example, a logistic regression, depending on covariates and on the last observed value of the repeated measurement process. It can also depend on the value that would have been observed if the drop-out process had not taken place.
- **Pattern-mixture model** (Wu and Carroll, 1988): a same underlying unobserved random effect specific to each individual is assumed for both the observed repeated measurement process and the drop-out process.

Many presentations were given on this topic illustrating the use of these models in practical cases, and studying their limitations:

- **Little**: *An assessment of pattern mixture models for handling potentially non-ignorable nonresponse.*
- **Molenberghs**: *Sensitivity analysis for pattern-mixture models.*
- **Carpenter**: *Dropout process in longitudinal studies - Can we restrict the scope of sensitivity analysis ?*
- **Dufouil**: *A comparison of approaches to the analysis of longitudinal studies with dropouts.*

Such models may be used in the case of genetic evaluation for milk production to deal with short lactations. At the moment, the lactation length is not taken into account in the calculation of genetic values.

2. Joint analysis of repeated measurements and time to key event

According to Liang and Zeger's course, this really corresponds to the next step for research on longitudinal data analysis. In fact, there are many cases occurring in medical studies, for instance, that correspond to this kind of problem. For example, in HIV studies, it would be interesting to study the causal effect of CD4 cell count and the occurrence of the disease. In the animal breeding area this kind of analysis can also be of interest, for example to model jointly milk production and survival time, to either study the causality of these two measures, or to take into account the health of the cow during the genetic evaluation for milk production.

The main presentations given on this subject where:

- **Liang and Zeger:** course.
- **Diggle:** *Some models for survival outcomes and associated longitudinal measurements.*

Their idea was to consider a classical longitudinal data model for the observed repeated measurement process, and a survival model for the time to key event. In order to have a joint analysis, they assumed an underlying bivariate process. However, it is usually assumed that the two underlying processes are proportional. This seems to be quite a restrictive assumption, and more research is needed on this subject in order to relax it and propose a more realistic model.

3. Repeated measurements for non-normally distributed traits

Many presentations concerned the analysis of observed binary repeated measurements. Issues were raised about estimation procedures since, in this case, likelihood does not have a close form.

- **Nelder and Lee:** *Random effects in generalized linear models.*
- **Renard:** *Pseudo-likelihood estimation in multilevel models with binary responses.*
- **Chang:** *Some modelling issues with correlated binary data.*
- **Gao:** *A shared random effect model approach for longitudinal binary data with non-ignorable missing data.*