Statistical Modelling 2014; **14**(**6**): i–ii

Guest Editorial

Special Issue: Statistical Methods in Oral Health Research

Oral health research is one discipline within medical research. So why is a special issue of the journal devoted to this particular medical discipline? The two guest editors of this special issue have been active in applying and developing statistical methods in dental research data for many years. Through their interdisciplinary collaboration, they are both convinced that oral health research provides a unique source for statistical developments. It is, however, not claimed that the statistical approaches in oral health research cannot be found in other application areas. But it is true that in oral health research, clinical research questions that matter quickly involve advanced statistical approaches. This is exemplified in the first contribution of this special issue. The two guest editors show in their contribution that a single rich dataset can be the source for numerous new statistical developments, dealing with non-standard distributions, misclassification and agreement issues, spatial associations in the mouth, interval censoring of events such as emergence and caries development, the complex multilevel and multivariate structure of dental data, etc. Clearly, such datasets may be available in other application areas; in oral health research though, they come by naturally. In the second contribution by Hsu *et al.*, the Wald test based on a quasi-likelihood approach is proposed to check for zero inflation and deflation in a dataset on dental caries experience. Their contribution is an example of modelling a non-standard distribution for counts, i.e., where there are possibly too many or too few zero occurrences compared to a standard distribution, e.g., a Poisson or a negative binomial distribution. Such a non-standard distribution occurs quite often with correlated counts and also with caries experience data, since teeth in the same mouth undergo the same environmental and dietary conditions which renders the counts correlated. Another source of correlation is the spatial position of the teeth such that caries development on one tooth can easily transfer to neighbouring teeth. Datta and Beck suggest using R-estimation, a robust estimation technique, to estimate marginal regression parameters for clustered data. Their approach allows that the cluster size is random and informative. Informative cluster sizes occur naturally in dental data with adult and elderly patients. For instance, the number of teeth in the mouth (cluster size) is itself the result of certain risk factors such as brushing and dietary behaviour. The robust estimation procedure is applied to a historical periodontal dataset. Parker *et al.* suggest in the third contribution to use a Bayesian spatial beta regression model, to fit proportions on the unit interval that can assume also the values of zero and one. The data are clinical attachment levels which are used to measure the periodontal disease status. In contrast to most spatial models, distance is not defined in this contribution on the geographical location of the subjects but within the subjects, and reflects the neighbouring structure

 10.1177/1471082X14535478

of teeth within a mouth. The fourth contribution handles on a tree-based method for variable selection for clustered survival data. The approach by Hallett *et al.* is an adaptation of the random forest approach of Breiman (2001) to build a prediction risk model. More specifically, the authors propose a permutation-based variable importance ranking method, based on the extremely randomized trees method for correlated censored data. Their approach is applied to tooth loss data. The final contribution of Eriksson *et al.* warns the user not to overinterpret the complex statistical output. Namely, the authors show that when marginally or conditionally specified models for clustered failure time data are applied, the parameter estimates can be quite biased when the unknown frailty variable is correlated with the variable of interest in a non-randomized setting. This is a cautionary note to statistical practitioners especially when interpreting the results of shared frailty models, which are believed to deal in an appropriate manner with the correlated survival times. Through simulations, which were inspired by a historical dental implant study, the authors showed that severe bias may occur if the frailty is associated with the observed covariates.

From this special issue, it must be clear that the complexity of the research questions in oral health research offers ample opportunities to develop novel methods for statistical modelling. But also oral health researchers can benefit from a better understanding of the problems encountered. Communication and collaboration between both research groups could lead to a considerable quality improvement in the field of oral health research and to novel statistical techniques that are useful in many other application areas.

The guest editors strongly believe that true multidisciplinary research is the way to go. That was their reason to initiate international dental-statistical meetings that deal with 'Methodological Issues in Oral Health Research'. These meetings have been organized since 2004 (Leuven), biannually in Ghent (2006), Gargagno (2008), Istanbul (2010), Graz (2012) and Adelaide (2014) each with a particular focus. This has created a network of oral health researchers aiming at (1) setup of activities to improve statistical education in the training of oral health research workers, (2) review of the statistical requirements in oral health research and (3) exchange of ideas about new developments in statistical research which can improve the outcome of oral health research.

This special issue has no intention to give a complete overview of this exciting multidisciplinary research area. Rather the selected papers show how rich and complex the statistical analysis can and needs to be. We hope you will enjoy the selection of contributions.

Emmanuel Lesaffre
Department of Biostatistics, Erasmus MC
Rotterdam, The Netherlands and
L-Biostat, KU
Leuven, Belgium

Dominique Declerck
Department of Oral Health Sciences, KU Leuven, Leuven, Belgium