

```
hglm2(y ~ crack0 + (1|specimen), family =
Gamma(link=log), disp = ~cycle, data = crack_
growth)
```

Each part of the model can also be checked using diagnostic plot functions included in the package. Furthermore, the conditional AIC (cAIC) can be computed which is equivalent to the deviance information criterion (DIC) applied in Bayesian statistics (see Lee & Noh, 2012). Thereby we can compare models and assess the assumptions of the model. It is therefore rather easy to compare a number of alternative models.

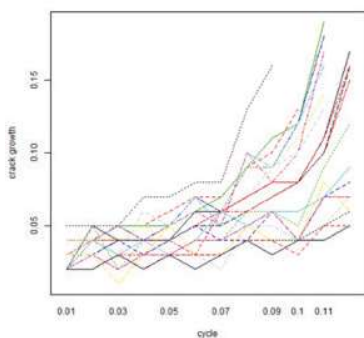


Figure 1: Crack lengths measured on a compact tension steel from 21 metallic specimens (Hudak et al. 1978).

HGLMs are further extended to double HGLMs (DHGLMs) by allowing random effects in the model for the dispersion parameters. In the previous model for the crack growth data specimen was included in the model for the mean. Now it is also possible to include specimen as a random effect in the model for the dispersion. In chapter 6 of Lee, Rönnegård & Noh (2017) these models are explored extensively and a model including specimen as random effect in the dispersion does indeed produce the lowest cAIC. HGLMs also allow random effects to fit various temporal-spatial or functional models. DHGLMs are further extended to multivariate DHGLMs (MDHGLMs), by allowing multi responses with various types. They are also extended to survival data, so that various frailty models and competing risks models can be fitted. These survival models can be

fitted jointly with other responses having MDHGLMs. Sparse variable selection and multiple testing can be also fitted via random effect models. More detailed explanations are available in monographs of Lee, Nelder and Pawitan (2017) and Ha, Jeong and Lee (2017).

A one-day course will be given in connection to IBC 2020 in Seoul. During this short course, organized by the conference, the participants will be introduced to the theory and examples given in the book Lee, Rönnegård & Noh (2017). The course includes several computer exercises using example data sets. The modelling will be executed using R codes with `hglm`, `dhglm`, `mdhglm`, `frailtyHL` and `jointdhglm` packages. Furthermore, a user-friendly interface with manuals for the whole ten chapters of the book will be provided putting more emphasis on the modelling rather than the R syntax. An overview of a wide range of applications, including Gaussian processes, factor analysis, survival analysis and variable selection models, based on the HGLM framework will be provided together with ongoing future developments.

References

- Ha, I.D., Jeong, J. & Lee, Y. (2017) Statistical modelling of survival data with random effects. Springer. ISBN 9789811065552.
- Lee, Y., Nelder, J.A. & Pawitan, Y. (2017) Generalized linear models with random effects. Chapman and Hall/ CRC. ISBN 9781498720618.
- Lee, Y., & Noh, M. (2012). Modelling random effect variance with double hierarchical generalized linear models. *Statistical Modelling*, 12(6), 487–502. DOI: 10.1177/1471082X12460132
- Lee, Y., Rönnegård, L. & Noh, M. (2017) *Data Analysis Using Hierarchical Generalized Linear Models with R*. Chapman and Hall/CRC. ISBN 9781138627826.

STrengthening Analytical Thinking for Observational Studies (STRATOS)

Introducing the Topic Group on Evaluating Diagnostic Tests and Prediction Models (TG6)

Ben Van Calster^{1,2}, Ewout W Steyerberg², on behalf of STRATOS TG6

¹ KU Leuven, Department of Development and Regeneration, Leuven, Belgium

² Department of Biomedical Data Sciences, Leiden University Medical Center, the Netherlands

Previous issues of the Bulletin introduced the STRATOS Initiative and activities of its various Topic Groups. In this note, we introduce the members and activities of TG6 of the STRATOS initiative, which is concerned with Evaluating Diagnostic Tests and Prediction Models. The chairs of TG6 are Ewout Steyerberg and Ben Van Calster.

The members of TG6 are renowned international researchers in the field of diagnostic and prognostic research. In alphabetical order: Patrick Bossuyt (Amsterdam University Medical Center), Gary Collins (University of Oxford), Petra Macaskill (University of Sydney), David McLernon (University of Aberdeen), Carl Moons (Utrecht University), Maarten van Smeden (Leiden University Medical Center), and Andrew Vickers (Memorial Sloan Kettering Cancer Center). We have natural links with all other topic groups in the STRATOS initiative, and specifically with TG2 (Selection of Variables and Functional Forms in Multivariable Analysis) and TG8 (Survival Analysis).

TG6 aims to produce guidance on methods for the development and evaluation of diagnostic tests and prediction models. Recent years have seen a spectacular rise in the number of assays, imaging

Region News

Australasian Region (AR)

<http://www.biometricsociety.org.au/about.html>

E. A. Cornish Award Awardees 2019

Established in 2011, the E. A. Cornish award is presented by the Australasian Region in recognition of a member's considerable contribution to biometry and the society. The award is named after Alf Cornish, the region's first Vice President and IBS President from 1956 to 1957. At the [2019 IBS-AR](#) conference dinner in Adelaide the two most recent awardees, Emlyn Williams and David Baird, were named. Congratulations!

[Photos](#) and [videos](#) from the award ceremony are available on the IBS-AR flickr page.

Citation for **Emlyn Williams** written by Kaye Basford and Mario D'Antuono.

Emlyn's efforts and productive outcomes have provided statisticians and non-statisticians, researchers and practitioners (from academia and industry) with invaluable assistance in efficient experimental design and the analysis of data collected from an extensive range of experiments. During his illustrious career, Emlyn has:

- Developed the methodology for the "alpha" class of designs for use in field, glasshouse and laboratory experiments and subsequently further developed latinized, row-column, spatial and p-rep designs.
- Developed and promulgated three software packages to assist users with research and practical application of these designs.
- Provided training and education nationally and internationally on experimental design and analysis.
- Distinguished himself as a leader, building and mentoring a team of statisticians at the Australian National University (ANU) which was responsible for providing support to all departments and PhD students.
- Continued to contribute to the advancement of biometry internationally following his retirement in 2013. Emlyn is currently an Emeritus Fellow at ANU and a co-editor on the forthcoming special issue in JABES on "Recent advances in the design and analysis of experiments and observational studies in agriculture".

Emlyn has been a prolific writer, having published 3 books, 8 book chapters, 138 journal articles, and manuals for various versions of 3 software packages, as well as editing 39 proceedings. While CycDesign is probably the most well-known of his software packages (with licenses in 18 countries), DataPlus (an effective interpretive system currently used in 31 countries) and OutMap (software to produce genetic linkage maps) are also widely used. He has collaborated extensively nationally and internationally, particularly with Nye John and more recently with Hans-Peter Piepho. Emlyn was an invited speaker to the IBC in Barcelona in 2018.

Emlyn's role as a statistical consultant (both within CSIRO and ANU) has meant that he has always been user focused with excellent communication skills. His ability to develop innovative statistical solutions

modalities, models and other tests that have been proposed. This is likely related to the recent popularity of flexible algorithms, including machine learning and artificial intelligence. However, reviews have indicated that many model development studies use suboptimal methods, that many lab tests and models are not properly validated, and that validation studies often do not address all relevant issues. In addition, development and validation studies for prediction models are poorly reported, which limits their usability. Some of the members of TG6 have been active in related initiatives like the EQUATOR network (<https://www.equator-network.org/>) and published reporting guidelines for diagnosis (STARD: Standards for reporting diagnostic accuracy studies) and prognosis (TRIPOD: Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis).

Members of the TG met in 2016 at the first general STRATOS meeting in Banff (Canada), followed by the 2018 meeting of three TGs (2,6,7) in Leiden (the Netherlands, <https://www.lorentzcenter.nl/ic/web/2018/1017/report.pdf>), and another great Banff meeting in 2019 (<http://www.birs.ca/events/2020>, including viewing of a Grizzly bear). Several important topics were outlined on which the TG could focus in further work.

Recently, the TG has produced a manuscript that gives nontechnical guidance on how to deal with risk groups and risk thresholds in the context of prediction models ^[1]. This work focuses on the importance of the clinical consequences that are associated with a chosen risk threshold. We also discuss that clinical consequences are not set in stone but tend to vary across countries, regions, and individuals.

A second nontechnical paper has been produced, in which we explain and illustrate the evaluation of risk calibration when externally validating a model ^[2]. We label calibration the 'Achilles heel of prediction', since heterogeneity between development and validation settings coupled with small sample sizes can destroy clinical usefulness. Both papers are connected: if a model is miscalibrated, which means that risk estimates are incorrect, the performance of a model at a chosen risk threshold is affected.

More topics will be addressed in the near future in the exciting area of test and prediction model evaluation, including considerations on sample size for model development and validation studies. Moreover, we are preparing an overview of sensible approaches to the evaluation of performance in survival models in collaboration with TG8. We hope to contribute to bridging the gap between theory and practice in this area of medical research.

1. Wynants L, van Smeden M, McLernon DJ, Timmerman D, Steyerberg EW, Van Calster B on behalf of the Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. Three myths about risk thresholds for prediction models. *BMC Med.* 2019 Oct 25;17(1):192. doi: 10.1186/s12916-019-1425-3.
2. Van Calster B, McLernon DJ, Van Smeden M, Wynants L, Steyerberg EW; on behalf of the Topic Group 6 of the STRATOS initiative. Calibration: the Achilles heel of predictive analytics. *BMC Med* 2019; 17:230. doi: 10.1186/s12916-019-1466-7, in press.