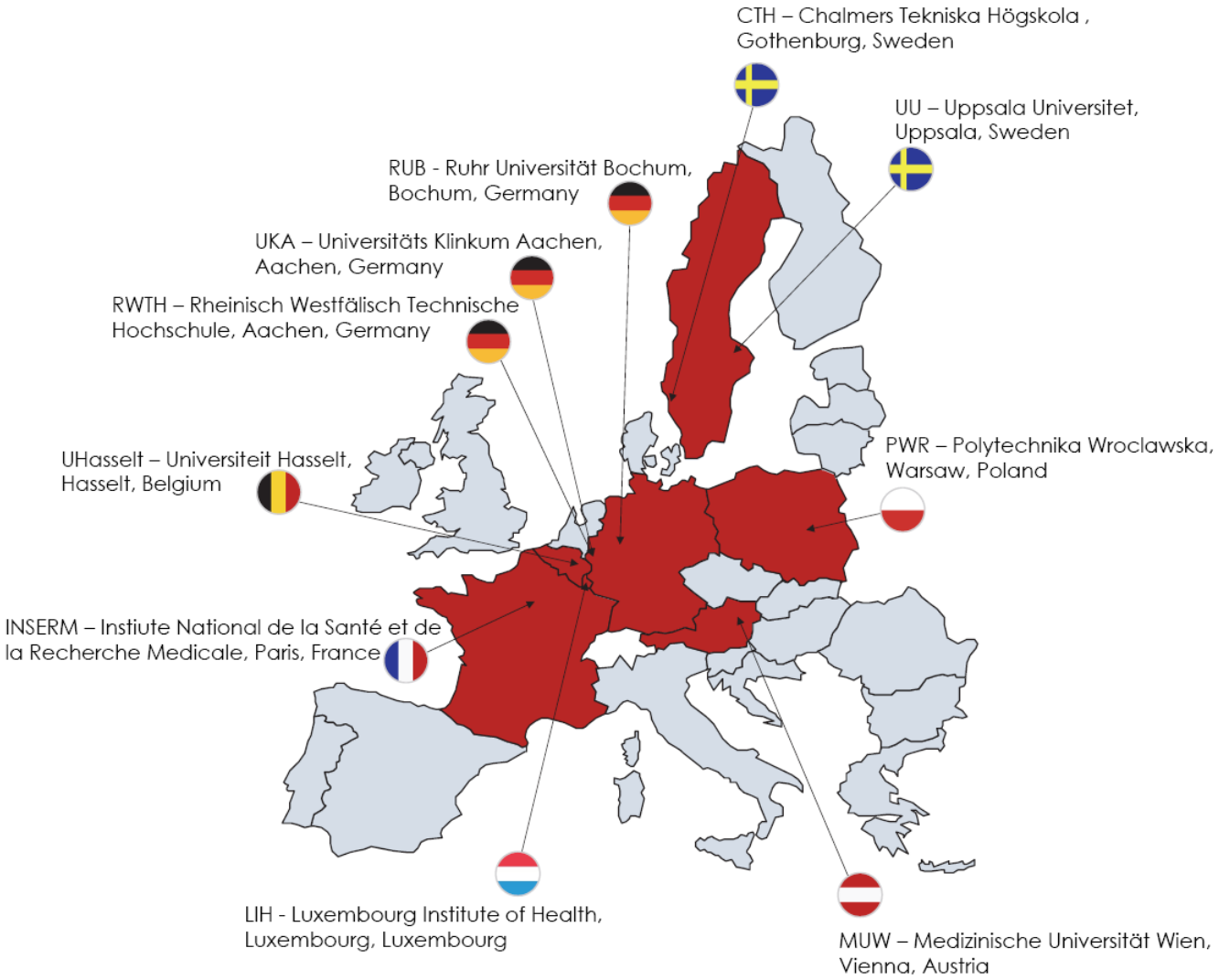
IDeAl Newsletter Issue 5

June 2016



Impressum

Project duration:	1. November 2013 -31. October 2016
Project coordinator:	Ralf-Dieter Hilgers University Hospital Aachen - Department of Medical Statistics
Project webpage:	http://www.ideal.rwth-aachen.de/
Social networks:	Twitter @ https://twitter.com/ideal_fp7
	LinkedIn @ https://www.linkedin.com/groups/IDEAL-FP7-Project-6556030
Newsletter Editors:	Gerald Hlavin (Medical University of Vienna) David Schindler, Diane Uschner (University of Aachen)

Highlight

Joint view of the three EU-FP7 funded projects on small populations has been published in the Orphanet Journal of Rare Diseases:

Hilgers, R. D., Roes, K., and Stallard, N. '**Directions for new developments on statistical design and analysis of small population group trials.**' Orphanet Journal of Rare Diseases. 2016.

Background

Most statistical design and analysis methods for clinical trials have been developed and evaluated where at least several hundreds of patients could be recruited. These methods may not be suitable to evaluate therapies if the sample size is unavoidably small, which is usually termed by small populations. The specific sample size cut off, where the standard methods fail, needs to be investigated. In this paper, the authors present their view on new developments for design and analysis of clinical trials in small population groups, where conventional statistical methods may be inappropriate, e.g., because of lack of power or poor adherence to asymptotic approximations due to sample size restrictions.

Method

Following the EMA/CHMP guideline on clinical trials in small populations, we consider directions for new developments in the area of statistical methodology for design and analysis of small population clinical trials. We relate the findings to the research activities of three projects, Asterix, IDeAl, and InSPiRe, which have received funding since 2013 within the FP7-HEALTH-2013-INNOVATION-1 framework of the EU. As not all aspects of the wide research area of small population clinical trials can be addressed, we focus on areas where we feel advances are needed and feasible.

Results

The general framework of the EMA/CHMP guideline on small population clinical trials stimulates a number of research areas. These serve as the basis for the three projects, Asterix, IDeAl, and InSPiRe, which use various approaches to develop new statistical methodology for design and analysis of small population clinical trials. Small population clinical trials refer to trials with a limited number of patients. Small populations may result from rare diseases or specific subtypes of more common diseases. New statistical methodology needs to be tailored to these specific situations.

Conclusion

The main results from the three projects will constitute a useful toolbox for improved design and analysis of small population clinical trials. They address various challenges presented by the EMA/CHMP guideline as well as recent discussions about extrapolation. There is a need for involvement of the patients' perspective in the planning and conduct of small population clinical trials for a successful therapy evaluation.

Latest IDeAl News posted on the website

February

- Prof Alan Agresti visted IDeAl Group at UKA

  Alan Agresti visited the IDEAL group for one week in February 2016 funded by an ERS grant of RWTH-Aachen University and became Theodore von Kármán Fellow.
- Workshop on Population Optimum Design of Experiments

  The 11th Workshop on Population Optimum Design of Experiments, PODE 2016, took place on Monday, 20 June 2016, in Uppsala, Sweden.
- Rare Disease Day 2016

  The Rare Disease Day raises awareness for rare diseases and their impact on patients' lives. Several events take place around the globe, all coordinated by EURORDIS.

March

- Rare Disease Taskforce meets in London

  On March 3, 2016, in London, UK, a Workshop was held at the EMA to discuss actions to reach agreement between the different stakeholders on appropriate small population studies. Under the roof of IRDiRC, a Task Force was founded to advance discussions on ways to optimise and improve commonly adopted approaches and to reach agreement between the different stakeholders on appropriate small population studies.

  Figure 1: IDEAL at the small group clinical trial workshop (IRDiRC) held at the EMA in London

  In this meeting, all three EU-funded projects on small population studies were involved:

  - IDEAL

    Ralf-Dieter Hilgers (IDeAl; RWTH Aachen, Germany)

    Mats Karlsson (Uppsala University, Sweden)

    Carl-Fredrik Burman (Astra Zeneca; Chalmers University, Sweden)

    Stephen Senn (Luxembourg Institute of Health, Luxembourg)

Franz König (Medical University Vienna, Austria)

Geert Molenberghs (KU Leuven, Belgium)

- IDEAL EAB

  Gérard Nguyen (Rett Syndrome Europe, France)

  Paolo Baroldi (Vanda Pharmaceuticals, USA)

  Frank Bretz (Novartis, Switzerland)

- ASTERIX

  Kit Roes (ASTERIX; UMC Utrecht, the Netherlands)

  Ferran Torres (Universitat Autònoma de Barcelona, Spain)

  Martin Posch (Medical University Vienna, Austria)

- INSPIRE

  Nigel Stallard (InSPiRe; University of Warwick, UK)

  Simon Day (Clinical Trials Consulting & Training Limited, USA)

  Tim Friede (University of Goettingen, Germany)

  Sarah Zohar (INSERM, France)

Figure 2: IRDiRC Task-Force members at the workshop on small population clinical trials

- Tutorial on Adaptive Designs for Confirmatory Clinical Trials

  On behalf of the IDEAL consortium Franz König will give an invited tutorial on „Adaptive Designs for Confirmatory Clinical Trials" at the ENAR 2016 Spring Meeting in Austin, Texas, on March 8. The meeting is organized by Eastern North American Region (ENAR) of the International Biometric Society (IBS).

- IDeAl researchers present latest results at DagStat

  Several researchers connected to the IDeAl group present their research at the Fourth Joint Statistical Meeting of the Deutsche Arbeitsgemeinschaft Statistik "Statistics under one Umbrella" that takes place at Georg-August University Göttingen this week (14th- 18th of March).

Figure 3: Young researchers of IDEAL and other small population projects

April

- IDeAl research spreads overseas

  IDeAl senior researcher Nicole Heussen and IDeAl coordinator Ralf-Dieter Hilgers disseminated IDeAl methodology to stakeholders from regulators to academia in the United States.

- IDeAl talk at FDA

  Ralf-Dieter Hilgers and Nicole Heussen give a talk at the U.S. Food and Drug Administration on 2016 May 6 titled "Does Randomization protect against bias? What can be done to improve the level of clinical evidence of effectiveness?"

Figure 4: IDeAl at GMU in Fairfax, Virginia

- Ralf-Dieter Hilgers talks about randomization at GMU

  On Friday, 22nd of April, Ralf-Dieter Hilgers gave a talk at George Mason University in Fairfax, Virginia, USA on the choice of randomization procedures.

- Talk of Nicole Heussen about "Strategies for dealing with missing data in randomization tests" on April 29th. at the Clifton Bailey Seminar Series Spring 2016, George Mason University in Fairfax, Virginia.

May

- IDeAl at Personalised Medicine Conference 2016
  Ralf-Dieter Hilgers and Nicole Heussen represent IDeAl at the Conference of the PerMed initiative in Brussels on 1-2 June. The Personalised Medicine Conference 2016 will explore personalised medicine through a research policy lens.

June

- IDeAl at mODa11
  Ralf-Dieter Hilgers (WP2), David Schindler (WP2) and Kirsten Schorning (WP3) represent IDeAl at the conference mODa11 in Hamminkeln-Dingden on 12-17 June.

Figure 5: Kirsten Schorning, David Schindler, Ralf-Dieter Hilgers discuss recent advances in optimum experimental design with EAB members Bill Rosenberger and Rosemary Bailey.

New Results

Research Articles (Peer-Reviewed)

- Alonso Abad, A., Van der Elst, W., Molenberghs, G., Burzykowski, T., & Buyse, M. On the relationship between the causal-inference and meta-analytic paradigms for the validation of surrogate endpoints. *Biometrics*, 71, 15-24, 2015.
- Senn, S. Mastering variation: variance components and personalised medicine. *Statistics in Medicine*, 35, 2, 966–977. 2016.
- Hecksteden, A., Kraushaar, J., Scharhag-Rosenberger, F., Theisen, D., Senn, S. and Meyer, T. Individual response to exercise training – a statistical perspective. *Journal of Applied Physiology*.118, 2, 1450-1459. 2016.
- Magirr, D., Jaki, T., Koenig, F., and Posch, M. Sample Size Reassessment and Hypothesis Testing in Adaptive Survival Trials. *PLoS ONE*, 11, e0146465. 2016.
- Lee, S., Brzyski, D., Bogdan M. Fast Saddle-Point Algorithm for Generalized Dantzig Selector and FDR Control with the Ordered l1-Norm. *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, JMLR:W&CP vol.51, 780–789, 2016.

- Riviere, M.-K., Ueckert, S., Mentré, F. A MCMC-method for the evaluation of the Fisher information matrix for nonlinear mixed effect models. *Biostatistics*, doi:10.1093/biostatistics/kxw020, 2016.
- Bauer, P., Bretz, F., Dragalin, V., Koenig, F., and Wassmer, G. Authors' response to comments ("Twenty-five years of confirmatory adaptive designs: opportunities and pitfalls"). *Statistics in Medicine,* 35, 364-367. 2016.
- Hilgers, R. D., Roes, K., and Stallard, N. Directions for new developments on statistical design and analysis of small population group trials. *Orphanet Journal of Rare Diseases*, 11:78, DOI: 10.1186/s13023-016-0464-5. 2016.
- Auffray, C., Balling, R., Barroso, I., Bencze, L., Benson, M., Bergeron, J., Bernal-Delgado, E., Blomberg, N., Bock, C., Conesa, A., Del Signore, S., Delogne, C., Devilee, P., Di Meglio, A., Eijkemans, R., Flicek, P., Graf, N:, Grimm, V., Guchelaar, H.-j., Guo, Y., Glynne Gut, I., Hanbury, A., Hanif, S., Hilgers, R.-D., Honrado, A., Hose, D., Houwing-Duistermaat, Jeanine, Hubbard, T., Janacek, S. H., Karanikas, H., Kievits, T., Kohler, M., Kremer, A., Lanfear, J., Lengauer, T., Maes, E., Meert, T., Muller, W., Nickel, D., Oledzki, P., Pedersen, B., Petkovic, M., Pliakos, K., Rattray, M., Redon i Mas, J., Schneider, R., Sengstag, T., Serra Picamal. X., Spek, W., Tome, M., Vaas, L., van Batenburg, O., Vandelaer, M., Varnai, P., Volloslada, P., Vzicaino J. A., Wubbe, J., Zanetti, G. Making sense of big data in health research: Towards an European Union action plan. *Genome Medicine*, 8:71,DOI: 10.1186/s13073-016-0323-y. 2016.

Further Articles and Pre-Prints

- Burman, C.-F. From optimal design theory to optimizing designs of clinical trials. Festschrift in Honor of Hans Nyquist on the Occasion of His 65th Birthday. (Ed: E. Fackle-Fornius), Stockholm Univ, 2015.
- Rübben, A., Hilgers, R.-D., Leverkus, M. Hedgehog Blockade for Basal Cell Carcinoma. Coming at a (Secondary Neoplastic) Price. *JAMA Dermatology*, 152(5):521-523, 2016.
- Jobjörnsson, S., Forster, M., Pertile, P., Burman, C.-F. Late-Stage Pharmaceutical R&D for Rare Diseases under Two-Stage Regulation. University of York, Department of Economics and Related Studies, Discussion Paper 15/16.
- Sobczyk, P., Bogdan M., Josse J. Bayesian dimensionality reduction with PCA using penalized semi-integrated likelihood. arXiv:1606.05333 [stat.ME]
- Brzyski, D., Peterson, C., Sobczyk, P., Candes, E., Bogdan, M., Sabatti, C. Controlling the rate of GWAS false discoveries. bioRxiv 058230; doi: http://dx.doi.org/10.1101/058230

Presentations

- "Does Randomization protect against bias?". R.-D. Hilgers and N. Heussen. 2016 May 6. FDA, Silver Spring, Maryland, USA.
- "A little bit me, a little bit you: N of 1 trials, random effects and shrinkage estimators" (invited lecture). S. Senn. 2016 May 2. Trends and Innovations in Clinical Trial Statistics, Durham, North Carolina, USA.
- "Strategies for dealing with missing data in randomization tests" (invited talk). N. Heussen. 2016 April 29. Clifton Bailey Seminar Series Spring 2016, George Mason University, Fairfax, Virginia, USA.
- "P-values? The problem is not what you think" (key note invited talk). S. Senn. 2016 April 22. FLAMES Annual Meeting 2016, Antwerp, Belgium.
- "Aspects for the scientific evaluation of randomization procedures in small clinical trials" (invited talk). R.-D. Hilgers. 2016 April 22. Clifton Bailey Seminar Series Spring 2016, George Mason University, Fairfax, Virginia, USA.
- "The Challenge of Small Data" (invited talk). S. Senn. 2016 March 24. Luxembourg Analytics Summit. Hamm, Luxembourg.
- "Validating the similarity of regresion curves with regard to small samples". K. Möllenhoff, H. Dette. 2016 March 16. DAGStat 2016, Göttingen, Germany.
- "The impact of bias on different randomization procedures". D. Schindler, N. Heussen, D. Uschner, R.-D. Hilgers. 2016 March 15. DAGStat 2016, Göttingen, Germany.
- "Optimizing trial designs for targeted therapies – A decision theoretic approach comparing sponsor and public health perspectives". T. Ondra, S. Jobjörnsson. 2016 February 11. Seminar der Wiener Biometrische Sektion der Internationalen Biometrischen Gesellschaft Region Österreich – Schweiz. Medical University of Vienna, Austria.
- "Late-Stage Pharmaceutical R&D for Rare Diseases under Two-Stage Regulation". S. Jobjörnsson, M. Forster (presenter), P. Pertile, C.-F. Burman. 2016 January 18. York Economics Workshop, York, Great Britain.
- "Can we optimize rare disease trials?" (invided talk). C.-F. Burman. 2016 April 25-27. DIA/FDA Statistics Forum, Bethesda, Maryland, USA.
- "How to adjust for multiplicity if you must: The Casino approach" (invided talk). C.-F. Burman. 2016 April 20. Health Metrics, Göteborg University, Sweden.
- "From Optimal Design Theory to Optimizing Designs of Clinical Trials" (invited talk). C.-F. Burman. 2015 December 9. Symposium (celebrating Hans Nyquist's 65 years), Stockholm University, Sweden.

- “Optimizing the biomarker subpopulation strategy in late stage clinical development” (invited talk). C.-F. Burman. 2015 June 14-18. DIA 51st Annual Meeting, Washington DC, USA.
- “Science, logic and patients. On inference in small disease populations” (invited session). C.-F. Burman. 2015 September 2-5. MCP, Hyderabad, India.
- "Late-Stage Pharmaceutical R&D for Rare Diseases under Two-Stage Regulation". S. Jobjörnsson, M. Forster, P. Pertile (presenter), C.-F. Burman. 2015 October 15-16. AEIS, Alghero, Italy.
- “Stakeholder perspectives on drug development decision making” (invited talk). C.-F. Burman. 2014 November 12. Biometric Colloquium, Hannover, Germany.
- “A Bayesian model for the selection of sample size in clinical trials”. S. Jobjörnsson. 2014 June 2-6. Nordstat, Turku, Finland.
- “Optimal designs for comparing curves”. H. Dette. 2015 December 13. University of London, UK.
- “Validating the similarity of regression curves with regard to small sample sizes”. K. Möllenhoff, H. Dette. 2016 March 14-18. DAGStat, Göttingen, Germany.
- “Pseudo-likelihood and Split-sample Methods in Small and Very Large Studies.” G. Molenberghs. 2016 May 23. Conference of the Brazilian Region of the International Biometric Society, Salvador, Brazil.
- “Mastering Variation” (invited lecture). S. Senn. 2016 May 5. Biogen, Cambridge MA, USA.
- “The Challenge of ‘Small Data’ – Rare Diseases and Ways to Study them”. S.Senn. 2016 May 5. Biogen, Cambridge MA, USA.
- “Computation of the Fisher information matrix for discrete nonlinear mixed effect models”. S. Ueckert, F. Mentré. 2015 Dec. 12-14. 8th International Conference in Computational and Methodological Statistics, London, UK.
- “Evidence, Eminence and Extrapolation - the background ” (invited talk). P. Bauer. 2016 May 3. Novartis Basel, CH.
- “Adaptive Designs for Confirmatory Clinical Trials” (invited tutorial). F. König. 2016 March 8. ENAR 2016 Spring Meeting, Austin, TX, USA.
- “Regulatory and methodological issues in adaptive designs for confirmatory trials” (invited talk). F. König and M. Posch. 2016 May 4. BBS Seminar Adaptive Designs, Basel, CH
- “An extrapolation framework to specify requirements for drug development in children” (invited talk). F. König. 2016 May 3. Novartis Basel, CH.
- “An extrapolation framework to specify requirements for drug development in children“ (invited talk). F. König. 2016 May 25. PSI Conference, Berlin, Germany.

Poster Presentations

- "Fast Saddle-Point Algorithm for Generalized Dantzig Selector and FDR Control with the Ordered l1-Norm". S. Lee, D. Brzyski, M. Bogdan. 19th International Conference on Artificial Intelligence and Statistics, Cadiz, Spain.
- "Marketing authorisation of orphan medicines in Europe 2000-2013: a 13-year experience". M. P. Hofer, H. Hedman, S. Tsigkos, T. Vetter, M. Mavris, J. Llinares Garcia, A. Elsaesser, M. Posch, F. König, S. Vamvakas, J. Regnström, S. Aarum. 2016 May 26-28. ECRD 2016: The European Conference on Rare Diseases & Orphan Products, Edinburgh, UK.
- „Minimierungsmethode bei binärem Outcome und kleiner Fallzahl." C. Fitzner, L. A. Dávila, R.-D. Hilgers.2016 March 14-18. DAGStat, Göttingen, Germany.

Short Courses and Workshops

- "Tutorial on Adaptive Designs for Confirmatory Clinical Trials". F. König. 2016 March 8. ENAR 2016 Spring Meeting. Austin, Texas, USA.
- "Statistical Issues in Drug Development". S. Senn. 2016 May 10-11. Luxembourg Institute of Health, Luxembourg.
- "Analysing and planning n-of-1 trials". S. Senn. 2016 April 28. Vanda Pharmaceuticals, Washington DC, USA.
- "Pseudo-likelihood and Split-sample Methods in Small and Very Large Studies". G. Molenberghs. 2016 January 13. London School of Hygiene and Tropical Medicine, London, UK.
- Summerschool (2 days) on "Group Sequential and Adaptive Study Designs", organized by Universität Salzburg, International Biometric Society (German Region and Region Austria-Swiss), Austrian Statistical Society. F. König, W. Brannath, M. Kieser and G. Rauch. 2016 June 29 – July 1. Strobl am Wolfgangsee, Austria.

Statistical Software

- R Package "bdpopt"(Sebastian Jobjörnsson): Optimisation of Bayesian Decision Problems
    - Optimisation of the expected utility in single-stage and multi-stage Bayesian decision problems. The expected utility is estimated by simulation. For single-stage problems, JAGS is used to draw MCMC samples.
- R Package "grpSLOPE" (Alexej Gossmann [Maintainer]): Group Sorted L1 Penalized Estimation
    - Group SLOPE is a penalized linear regression method that is used for adaptive selection of groups of significant predictors in a high-dimensional linear model. The Group SLOPE method can control the (group) false discovery rate at a user-specified level (i.e., control the expected proportion of irrelevant among all selected groups of predictors).
    - Authors: Alexej Gossmann, Damian Brzyski, Weijie Su, Malgorzata Bogdan, Ewout van den Berg, Emmanuel Candes (A part of the optimization code was obtained from http://statweb.stanford.edu/~candes/SortedL1/software.html under GNU GPL-3)

Abstracts of Peer-Reviewed Articles

Alonso Abad, A., Van der Elst, W., Molenberghs, G., Burzykowski, T., & Buyse, M. **'On the relationship between the causal-inference and meta-analytic paradigms for the validation of surrogate endpoints.'** *Biometrics*. 2015 Mar;71(1):15-24. doi: 10.1111/biom.12245.

The increasing cost of drug development has raised the demand for surrogate endpoints when evaluating new drugs in clinical trials. However, over the years, it has become clear that surrogate endpoints need to be statistically evaluated and deemed valid, before they can be used as substitutes of "true" endpoints in clinical studies. Nowadays, two paradigms, based on causal-inference and meta-analysis, dominate the scene. Nonetheless, although the literature emanating from these paradigms is wide, till now the relationship between them has largely been left unexplored. In the present work, we discuss the conceptual framework underlying both approaches and study the relationship between them using theoretical elements and the analysis of a real case study. Furthermore, we show that the meta-analytic approach can be embedded within a causal-inference framework on the one hand and that it can be heuristically justified why surrogate endpoints successfully evaluated using this approach will often be appealing from a causal-inference perspective as well, on the other. A newly developed and user friendly R package Surrogate is provided to carry out the evaluation exercise.

Senn, S. **'Mastering variation: variance components and personalised medicine.'** *Stat Med.* 2016 Mar 30;35(7):966-77. doi: 10.1002/sim.6739.

Various sources of variation in observed response in clinical trials and clinical practice are considered, and ways in which the corresponding components of variation might be estimated are discussed. Although the issues have been generally well-covered in the statistical literature, they seem to be poorly understood in the medical literature and even the statistical literature occasionally shows some confusion. To increase understanding and communication, some simple graphical approaches to illustrating issues are proposed. It is also suggested that reducing variation in medical practice might make as big a contribution to improving health outcome as personalising its delivery according to the patient. It is concluded that the common belief that there is a strong personal element in response to treatment is not based on sound statistical evidence.

Hecksteden, A., Kraushaar, J., Scharhag-Rosenberger, F., Theisen, D., Senn, S. and Meyer, T. **'Individual response to exercise training - a statistical perspective.'** *J Appl Physiol*. 2015 Jun 15;118(12):1450-9. doi: 10.1152/japplphysiol.00714.2014.

In the era of personalized medicine, interindividual differences in the magnitude of response to an exercise training program (subject-by-training interaction; "individual response") have received increasing scientific interest. However, standard approaches for quantification and prediction remain to be established, probably due to the specific considerations associated with interactive effects, in particular on the individual level, compared with the prevailing investigation of main effects. Regarding the quantification of subject-by-training interaction in terms of variance components, confounding sources of variability have to be considered. Clearly, measurement error limits the accuracy of response estimates and thereby contributes to variation. This problem is of particular importance for analyses on the individual level, because a low signal-to-noise ratio may not be compensated by increasing sample size (1 case). Moreover, within-subject variation in training efficacy may contribute to gross response variability. This largely unstudied source of variation may not be disclosed by comparison to a control group but calls for repeated interventions. A second critical point concerns the prediction of response. There is little doubt that exercise training response is influenced by a multitude of determinants. Moreover, indications of interaction between influencing factors of training efficacy lead to the hypothesis that optimal predictive accuracy may be attained using an interactive rather than additive approach. Taken together, aiming at conclusive inference and optimal predictive accuracy in the investigation of subject-by-training interaction entails specific requirements that are deducibly based on statistical principles but beset with many practical difficulties. Therefore, pragmatic alternatives are warranted.

Magirr, D., Jaki, T., Koenig, F., and Posch, M. **'Sample Size Reassessment and Hypothesis Testing in Adaptive Survival Trials'** *PLoS One*. 2016; 11(2): e0146465.

Mid-study design modifications are becoming increasingly accepted in confirmatory clinical trials, so long as appropriate methods are applied such that error rates are controlled. It is therefore unfortunate that the important case of time-to-event endpoints is not easily handled by the standard theory. We analyze current methods that allow design modifications to be based on the full interim data, i.e., not only the observed event times but also secondary endpoint and safety data from patients who are yet to have an event. We show that the final test statistic may ignore a substantial subset of the observed event times. An alternative test incorporating all event times is found, where a conservative assumption must be made in order to guarantee type I error control. We examine the power of this approach using the example of a clinical trial comparing two cancer therapies.

Lee, S., Brzyski, D., Bogdan M. **'Fast Saddle-Point Algorithm for Generalized Dantzig Selector and FDR Control with the Ordered l1-Norm.'** Proceedings of the 19th International Conference on Artificial Intelligence and Statistics, JMLR:W&CP vol.51, 780–789, 2016.

In this paper we propose a primal-dual proximal extragradient algorithm to solve the generalized Dantzig selector (GDS) estimation problem, based on a new convex-concave saddle-point (SP) reformulation. Our new formulation makes it possible to adopt recent developments in saddle-point optimization, to achieve the optimal $O(1/k)$ rate of convergence. Compared to the optimal non-SP algorithms, ours do not require specification of sensitive parameters that affect algorithm performance or solution quality. We also provide a new analysis showing a possibility of local acceleration to achieve the rate of $O(1/k2)$ in special cases even without strong convexity or strong smoothness. As an application, we propose a GDS equipped with the ordered ℓ1-norm, showing its false discovery rate control properties in variable selection. Algorithm performance is compared between ours and other alternatives, including the linearized ADMM, Nesterov's smoothing, Nemirovski's mirror-prox, and the accelerated hybrid proximal extragradient techniques.

Abstracts of Articles in Peer-Reviewed Journals

Burman, C.-F. **'From optimal design theory to optimizing designs of clinical trials.'** Festschrift in Honor of Hans Nyquist on the Occasion of His 65th Birthday. (Ed: E. Fackle-Fornius), Stockholm Univ, 2015.

Optimal design theory is applicable to certain aspects of the design of clinical trials. In this article, we will discuss D-optimal designs for Emax models in particular. However, several important design features are outside the scope of classic optimal design theory. One example is optimization of the sample size. Solving this type of problems requires us to move from a narrow view of statistics to an appreciation of the design as part of a wider scientific context. This may be especially important when considering trials in rare diseases where few patients are available for trial inclusion, the cost is relatively large compared to potential drug sales and where much is at stakes for future patients and patients in the trial. A particularly challenging problem is that of programme optimisation, where a dose-finding trial is to be optimised, not based on a function of its Fisher information matrix, but based on the expected utility for the optimal design of the following confirmatory trial.

Jobjörnsson, S., Forster, M., Pertile, P., Burman, C.-F. **'Late-Stage Pharmaceutical R&D for Rare Diseases under Two-Stage Regulation.'** University of York, Department of Economics and Related Studies, Discussion Paper 15/16.

We present a model combining the two regulatory stages relevant to the approval of a new health technology: the authorisation of its commercialisation and the insurer's decision about whether to reimburse its cost. We show that the degree of uncertainty around the true value of the insurer's maximum willingness to pay for a unit increase in effectiveness has a non-monotonic impact on the price of the innovation, the firm's expected profit and the optimal sample size chosen for the clinical trial. A key result is that there exists a range of values of the uncertainty parameter over which a reduction in uncertainty benefits the firm, the

insurer and patients. We consider how different policy parameters may be used as incentive mechanisms, and the incentives to invest in R&D for marginal projects such as those targeting rare diseases. The model is calibrated using data on a new treatment for cystic fibrosis.

Sobczyk, P., Bogdan M., Josse J. **'Bayesian dimensionality reduction with PCA using penalized semi-integrated likelihood.'** arXiv:1606.05333 [stat.ME]

Our goal is to identify genes which regulate genetic pathways related to the disease development and response to the treatment. For this purpose we developed an iterative subspace clustering algorithm *varclust* which separates genes into pathways based on gene expression data. In this approach it is assumed that expressions of genes in one pathway can be represented as linear combinations of few regulating factors. For the sake of performance of this algorithm, it is essential to have a precise estimate of the number of these factors. In statistical terms, this reduces to the problem of estimating the number of principal components in PCA. In our context the problem is difficult, since the number of genes in one pathway can often be substantially larger than the number of patients for whom the data were collected. To solve this problem we consider approximate Bayesian criterion for a fixed effects PCA model: $\mathbf{X}_{n\times p} = \mathbf{M}_{n\times p} + E$, where **M** is assumed to be of low rank $k \leq \min(n,p)$ and $E = [\epsilon_{i,j}]_{n\times p}$ is matrix of i.i.d. errors, $\epsilon_{i,j} \sim \mathcal{N}(0, \sigma^2)$. In this context classical Bayesian Information Criterion (BIC) cannot be used, since the number of parameters increases both in *n* and *p*. In our PEnalized SEmi-integrated Likelihood (*PESEL*) method, likelihood is integrated out with respect to some prior, which allows to reduce the number of free parameters so it does not longer depend on *p* or on *n*. This allows to build model selection criteria when $n >> p$ or when $n << p$. The latter is a natural setting for small populations.

In the figure we present results of robustness study in which we compared *PESEL* to up-to-date methods. It proves robust compared to Generalized Cross Validation (GCV), at the cost of slightly worse performance for weak signals. Passemier's method is comparably robust, but inferior to both PESELs in terms of retaining true dimensionality in noisy data.

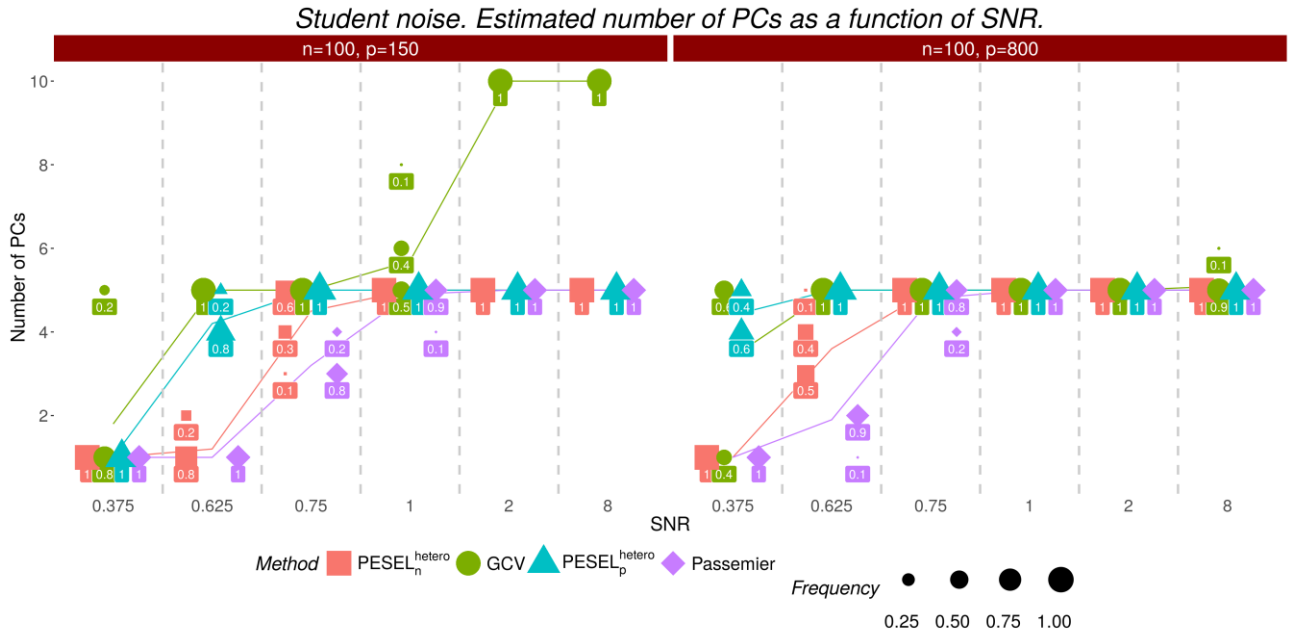


Figure 6: We generated data according to the fixed effects model, but with noise drawn from student distribution $\epsilon_{i,j} \sim t(df$ =3). True number of components k is 5. Lines represent mean estimated number of Principal Components, while the size of the symbols is proportional to the frequency of selection of a given number of PCs. Results are for number of variables p 150 and 800. SNR is the ratio between l2 norm of the signal to the variance of the noise. Number of observations n is constant and equal 100.

Brzyski, D., Peterson, C., Sobczyk, P., Candes, E., Bogdan, M., Sabatti, C. **'Controlling the rate of GWAS false discoveries'** bioRxiv 058230; doi: http://dx.doi.org/10.1101/058230

With the rise of both the number and the complexity of traits of interest, control of the false discovery rate (FDR) in genetic association studies has become an increasingly appealing and accepted target for multiple comparison adjustment. While a number of robust FDR controlling strategies exist, the nature of this error rate is intimately tied to the precise way in which discoveries are counted, and the performance of FDR controlling procedures is satisfactory only if there is a one-to-one correspondence between what scientists describe as unique discoveries and the number of rejected hypotheses. The presence of linkage disequilibrium between markers in genome-wide association studies (GWAS) often leads researchers to consider the signal associated to multiple neighboring SNPs as indicating the existence of a single genomic locus with possible influence on the phenotype. This a posteriori aggregation of rejected hypotheses results in inflation of the relevant FDR. We propose a novel approach to FDR control that is based on pre-screening to identify the level of resolution of distinct hypotheses. We show how FDR controlling strategies can be adapted to account for this initial selection both with theoretical results and simulations that mimic the dependence structure to be expected in GWAS. We demonstrate that our approach is versatile and useful when the data are analyzed using both tests based on single marker and multivariate regression. We provide an R package that allows practitioners to apply our procedure on standard GWAS format data, and illustrate its performance on lipid traits in the NFBC66 cohort study.

Riviere, M.-K., Ueckert, S., Mentré, F. **'A MCMC-method for the evaluation of the Fisher information matrix for nonlinear mixed effect models.'** Biostatistics, doi:10.1093/biostatistics/kxw020, 2016.

Non-linear mixed effect models (NLMEMs) are widely used for the analysis of longitudinal data. To design these studies, optimal design based on the expected Fisher information matrix (FIM) can be used instead of performing time-consuming clinical trial simulations. In recent years, estimation algorithms for NLMEMs have transitioned from linearization toward more exact higher-order methods. Optimal design, on the other hand, has mainly relied on first-order (FO) linearization to calculate the FIM. Although efficient in general, FO cannot be applied to complex non-linear models and with difficulty in studies with discrete data. We propose an approach to evaluate the expected FIM in NLMEMs for both discrete and continuous outcomes. We used Markov Chain Monte Carlo (MCMC) to integrate the derivatives of the log-likelihood over the random effects, and Monte Carlo to evaluate its expectation w.r.t. the observations. Our method was implemented in R using Stan, which efficiently draws MCMC samples and calculates partial derivatives of the log-likelihood. Evaluated on several examples, our approach showed good performance with relative standard errors (RSEs) close to those obtained by simulations. We studied the influence of the number of MC and MCMC samples and computed the uncertainty of the FIM evaluation. We also compared our approach to Adaptive Gaussian Quadrature, Laplace approximation, and FO. Our method is available in R-package *MIXFIM* and can be used to evaluate the FIM, its determinant with confidence intervals (CIs), and RSEs with CIs.

Auffray, C., Balling, R., Barroso, I., Bencze, L., Benson, M., Bergeron, J., Bernal-Delgado, E., Blomberg, N., Bock, C., Conesa, A., Del Signore, S., Delogne, C., Devilee, P., Di Meglio, A., Eijkemans, R., Flicek, P., Graf, N:, Grimm, V., Guchelaar, H.-j., Guo, Y., Glynne Gut, I., Hanbury, A., Hanif, S., Hilgers, R.-D., Honrado, A., Hose, D., Houwing-Duistermaat, Jeanine, Hubbard, T., Janacek, S. H., Karanikas, H., Kievits, T., Kohler, M., Kremer, A., Lanfear, J., Lengauer, T., Maes, E., Meert, T., Muller, W., Nickel, D., Oledzki, P., Pedersen, B., Petkovic, M., Pliakos, K., Rattray, M., Redon i Mas, J., Schneider, R., Sengstag, T., Serra Picamal. X., Spek, W., Tome, M., Vaas, L., van Batenburg, O., Vandelaer, M., Varnai, P., Volloslada, P., Vzicaino J. A., Wubbe, J., Zanetti, G. **'Making sense of big data in health research: Towards an European Union action plan.'** Genome Medicine, 8:71, doi:10.1186/s13073-016-0323-y. 2016.

Medicine and healthcare are undergoing profound changes. Whole-genome sequencing and high-resolution imaging technologies are key drivers of this rapid and crucial transformation. Technological innovation combined with automation and miniaturization has triggered an explosion in data production that will soon reach exabyte proportions. How are we going to deal with this exponential increase in data production? The potential of "big data" for improving health is enormous but, at the same time, we face a wide range of challenges to overcome urgently. Europe is very proud of its cultural diversity; however, exploitation of the data made available through advances in genomic medicine, imaging, and a wide range of mobile health applications or connected devices is hampered by numerous historical, technical, legal, and political barriers. European health systems and databases are diverse and fragmented. There is a lack of harmonization of data formats, processing, analysis, and data transfer, which leads to incompatibilities and lost opportunities. Legal frameworks for data sharing are evolving. Clinicians, researchers, and citizens need improved methods, tools, and training to generate, analyze, and query data effectively. Addressing these barriers will contribute to creating the European Single Market for health, which will improve health and healthcare for all Europeans.

The research leading to these results has received funding from the European Union Seventh Framework Programme under grant agreement n° 602552.