2016 PSI Conference

The Pullman Hotel, Berlin 22 - 25 May 2016

Programme of Abstracts

Promoting Statistical Insight and Collaboration in Drug Development

22 - 25 May 2016

Welcome Letter

On behalf of the PSI Conference Organising Committee, I'd like to welcome you to Berlin and The Pullman Hotel for the 39th Annual PSI Conference. PSI is excited to host its annual conference in Germany with a conference theme of Promoting Statistical Insight and Collaboration in Drug Development.

We are delighted to welcome our two keynote speakers who will be opening the conference on Monday and Tuesday: Paul Wicks, Vice President of Innovation at PatientsLikeMe and Prof. Dr. med. Stefan Lange, Deputy Director at IQWiG. This conference saw a record number of contributed abstracts submitted for oral presentation such that we had to increase the number of sessions to accommodate as many people as possible. Over the three days we can now look forward to three plenary sessions, 24 parallel sessions, with a total of more than 60 speakers.

The Conference App proved to be a huge success last year so it is available for download again this year and will keep you up to date with all the latest information, session plan, speaker abstracts and biographies.

I would like to invite you to take advantage of all of the opportunities this conference brings with it; in meeting with old colleagues and friends, making new associates, learning something new, and above all, having fun! This is my final year as Conference Chair (the time has flown!) and I couldn't do it without the constant support of a fantastic committee.

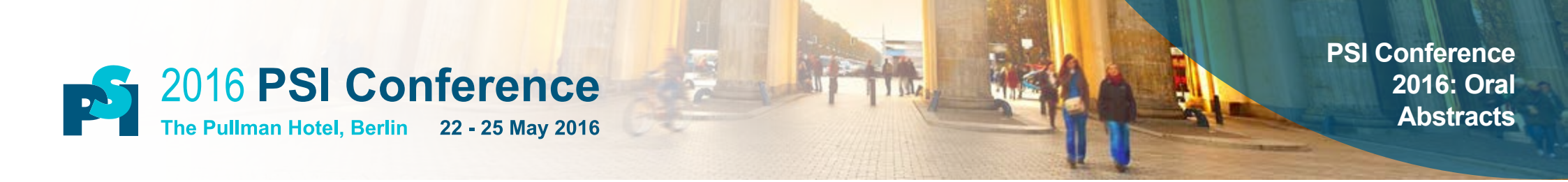
"I would like to invite you to take advantage of all of the opportunities this conference brings with it; in meeting with old colleagues and friends, making new associates, learning something new, and above all, having fun!"

The Scientific Committee currently comprises 18 statisticians (to see who, go to the Committee page on the App) working at various companies across the pharmaceutical industry who work hard all year, on a voluntary basis, to put together an agenda that will be relevant and interesting to such a wide audience. I hope you all agree that the programme looks fantastic and, like me, you will find it difficult to choose between the sessions. Thank you to everyone who has been involved in the organisation of the conference! I would also like to take the opportunity to thank all of our exhibitors and sponsors as we would be unable to run this event without your continuing support.

As usual, after the conference we will be contacting you with a link to the electronic feedback form. Your feedback is very important to us in planning future conferences and we especially welcome ideas for the future or ways to further improve the conference to make it a better experience for you.

I look forward to meeting as many of you as I can this year, and wish you all an enjoyable and successful conference.

Emma Jones, Veramed Limited
Conference Chair



PSI Conference 2016: Collated Oral Abstracts

Monday:

Opening Remarks

Mark Morris, PSI Chair, will welcome you to Berlin and formally open the 2016 PSI Conference.

Not Just Another Statistic: How Patients are Taking Control of Data and Research

Our keynote speaker Paul Wicks, Ph.D., Vice President of Innovation at PatientsLikeMe will open the conference.

Traditionally the role of patients in research has been to provide data and enrol in clinical trials as blinded participants. Today, however, the widespread availability of technology, education, and peer networks means that increasingly patients themselves have the tools and the support to begin exploring, generating, and testing their own data. Patient advocates are fighting to be heard too in how research funding is allocated, conducted, and disseminated. As patients seek to learn more about their conditions they are beset on all sides by the same statistical and inferential pitfalls that plague all of medical research, compounded by their potential lack of objectivity. But with the right support, and engagement from statistical experts, could patients and their carers be the greatest untapped resource medical research has ever seen?

Incorporating Real World Evidence with Randomised Controlled Trials in Network Meta-Analyses and Predicting Effectiveness

The work presented is from the IMI GetReal project:
http://www.imi-getreal.eu/

Speakers:

1. Orestis Efthimiou (University of Ioannina): "Methods for combining randomized and non-randomized evidence in a network meta-analysis"

Network meta-analyses (NMAs) are commonly employed in comparative effectiveness assessments for comparing medicines used in clinical practice and for assessing the efficacy and safety of a new medicine relative to existing therapies. Applications of NMA are often limited to the synthesis of evidence coming from randomized controlled trials (RCTs), while non-randomized evidence is often disregarded. Observational studies, however, convey valuable information about the effectiveness of interventions in real-life clinical practice and in recent years there has been a growing interest for methods to include them in the decision-making process. In this workshop we will discuss existing approaches and we will present new methods for incorporating non-randomized evidence in a NMA of RCTs, highlighting the advantages, limitations and key challenges of each approach. We will illustrate our methods using a network of pharmacological treatments for schizophrenia.

2. Thomas Debray (University Medical Centre, Utrecht): "Network meta-analysis using IPD - an illustration of its potential advantages"

With increasing access to individual patient data (IPD), it is easier to incorporate patient level covariates in a network meta-analysis (NMA). The different statistical models for how to incorporate IPD into an NMA and the challenges this brings will be presented and discussed.

3. Eva-Maria Didden (University of Berne, Switzerland): "Learning and Predicting Real-World Treatment Effect based on Randomized Trials and Observational Data: A case study on rheumatoid arthritis"

Real world evidence can be incorporated with RCT data to predict effectiveness. An approach for how to model and predict effectiveness using efficacy outcomes from RCTs and outcomes observed in clinical practice will be presented.

Utility of Translational Biomarkers in Clinical Development

The National Institutes of Health (NIH) defines translational research as the movement of discoveries in basic research to application at the clinical level. In this session we will hear examples, where different biomarkers explored through clinical development programs have been used to optimize patient selection and trial design. Please see the abstracts for more details and we hope to see you at this session.

Speakers:

1. Markus Lange (Novartis): "Analysis of clinical trial for biologics with a biomarker endpoint"

Biologics such as monoclonal antibodies are increasingly and successfully used for the treatment of many diseases. Unlike conventional small drug molecules, which are commonly given as tablets once daily, biologics are typically injected at much longer and often unequally spaced time intervals. Early phase clinical studies should provide information on the dose-time-response relationship, so that optimal regimens can be used in later stage clinical trials. In early phase clinical trials, typically biomarkers rather than clinical endpoints are measured, which then provide the basis for building dose-time-response models. We will discuss semi-mechanistic models that can also be used when patients receive different dosing regimens. We will consider a Bayesian approach for inference using the new software stan. The methodology is illustrated based on results from a clinical study with a monoclonal antibody, with tumor volume as the biomarker.

2. Nicola Voyle (King's College London):

Blood Metabolite Markers of Brain Amyloid-β Burden

It is thought that Alzheimer's Disease (AD) may be caused by the aggregation of amyloid-β (Aβ) in the brain and as such many anti-amyloid therapeutics are currently being tested in the quest to find a disease-modifying treatment for AD. It is desirable that trials of these therapeutics only recruit subjects with a high Aβ burden. Currently, Aβ burden is quantified through measurements in the cerebrospinal fluid (CSF) or via Positron Emission Tomography (PET) imaging. Both are invasive techniques that cause patient discomfort and can be expensive, and hence impractical, on a large scale. Consequently, there is high demand for a blood-based biomarker of elevated Aβ in the brain.

I will present results from the first study to investigate the association of blood metabolites with Aβ in the brain. We found a panel of 5 metabolite features that can predict high/low Aβ burden with 72% accuracy. I will also show how the inclusion of protein information further improves this accuracy.

If replicated in large, independent studies these metabolites and proteins could form the basis of a blood test with potential for enrichment of amyloid pathology in anti-Aβ trials.

3. Bianca Papi (OCS Consulting B.V.): "Penalized regression model to select a panel of saliva biomarkers from 17 adults affected by atopic dermatitis. A case-study on the application of LASSO models in small sample size."

Background: The primary objective of this study was to select a panel of saliva biomarkers correlated with disease severity in atopic dermatitis. Saliva has been collected from 17 atopic dermatitis patients treated with potent topical corticosteroids. Disease severity has been determined by LSS (Leicester Sign Score) and 73 biomarkers have been collected for each patient.

Methodology: Due to data's high-dimension, the small sample size and the high correlation among biomarkers, multivariate methods were not successful in reducing data dimension and they did not reach the goal of variable selection. Therefore, there was the need to apply statistical models that are more appropriate to study objective.

Stepwise regression analysis has been performed on the complete data set, in order to select the most correlated and informative panel of biomarkers. However, the Stepwise regression model had the problem of instability and over-fitting of subset selection by retaining redundant information and noisy variables. It provided biased estimators and a multicollinearity issue was present.

A penalized regression model (LASSO) has been proposed in order to avoid the over-fitting and multicollinearity issues. Penalized regression model aims, indeed, to parsimony by using a constraint on predictors. Since the overall magnitude of the coefficients is constrained, important predictors are included in the model, and less important predictors shrink, potentially to zero.

This methodology can help for automatic feature/variable selection and/or when dealing with highly correlated predictors, where standard regression will usually have regression coefficients that are 'too large" and create unbiased estimates.

Modelling and Simulation: Practice and Best Practice

The European Federation of Pharmaceutical Industries & Associations (EFPIA) working group on Model Informed Drug Discovery and Development (MID3) has just published a wide-ranging 90-page paper on the whys and hows of Best Practice. At the same time, the Board of PSI in December adopted a Best Practice document proposed by the PSI Modelling and Simulation Special Interest Group (SIG). The SIG Best Practice document could be regarded as a template for implementing the agreed Best Practice that is adumbrated in the MID3 document. In this interactive session, the two Best Practice initiatives will be described and agreed key requirements for best practice will be enumerated. The level of pre-specification necessary for Best Practice is a point that remains open and may vary – this and other points will be debated and explored via examples of Modelling and Simulation projects.

This session will be led by Michael O'Kelly (Quintiles), Alun Bedding (AstraZeneca), Chris Jennison (University of Bath) and Tom Parke (Berry Consultants).

PSI AGM

All PSI members are entitled to attend and speak at the Annual General Meeting (AGM).

To Proceed or Not to Proceed? Decision Criteria in Clinical Development

1. Peter Colman (UCB): "Trials, Tribulations and Errors – Deciding on Decision Criteria"

Today there are opportunities for statisticians to introduce quantitative decision criteria into many aspects of early phase clinical studies. We may implement model-based dose-escalation schemes, for individual subjects or groups of subjects, which focus on safety, markers of potential efficacy or both. Or we may lead the design of study level decision criteria, against which the success of the whole trial will be assessed. A willingness to accept increasing adaptation in early-phase clinical trials allows us to define decision criteria for dropping dose groups during the study. We consider the impact of the Bayesian paradigm on this mix and the challenges that we face in implementation.

2. Paul Frewer (AstraZeneca): "Decision Making in Early Clinical Development"

Throughout development, AstraZeneca uses a standard approach to set study decision criteria. This is key in Early Clinical Development where it is essential to make robust decisions quickly to enable compounds to progress or to be stopped. The approach uses a three outcome design and is based on the confidence of observing a result better than a Lower Reference Value (LRV) for a Go decision or a result worse than a Target Value (TV) for a No Go decision. If neither of these criteria are met, the outcome is in a "grey" zone, where teams need to use other information to move to a clear decision. When setting the criteria teams objectively set the LRV and TV based on the available evidence and agree the required confidence levels. They also assess the operating characteristics of the decision framework such that the probability of being in the "grey" zone is acceptable. Using this approach has enabled decision criteria in Early Clinical Development to be based on the same framework with the use of standard displays in agreeing these criteria with senior stakeholders. We are looking at extensions to this approach, to calculate the sample size, to incorporate interim analyses into the framework and to move it to a purely Bayesian approach.

3. Richard Vonk (Bayer): "Quantitative Decision Making – One Step Further"

Statistical sciences are currently moving into the focus of applied pharmaceutical research. The high costs and long duration of clinical development, paired with high levels of attrition, require the quantification of the risk when moving from early to late stage clinical development.

Further to the regulatory requirements, statistics and statistical thinking are integral parts of the internal decision making processes, particularly in early clinical development. This presentation concentrates on innovative statistical methods in different areas of early drug development that facilitate quantitative rather than qualitative decision making. We describe applications of (Bayesian) statistical techniques to improve decision making and decrease trial sizes. Furthermore, we explain how we implement this new way of thinking into our organization.

Focusing on Statistical Challenges in Oncology

1.Andy Stone (AstraZeneca): "Current and Future Statistical Challenges in Oncology"

As the knowledge of the biology of cancer has increased so has the investment of pharmaceutical companies and the availability of new treatments. The number of potential therapies, diseases and society's

desire to make important advances in treatment available to patients early creates challenges in the design of development programmes: if a new therapy receives accelerated approval on an intermediate endpoint how can we assess the effect on survival? ; if a promising agent is potentially effective in some, but not all types of cancer, how do we identify those without delaying development.

There are, and will continue to be methodological challenges too, for example: how to best describe the benefit of treatments, the best trial designs to identify the subgroup of patients who benefit, the use of intermediate endpoints, handling of missing data and multiplicity.

Statisticians will continue to play a crucial role oncology, and this presentation will share ideas on how we contribute to making meaningful new medicines available to patients.

2. Tina van der Horst (Roche): "Adaptive seamless phase II/III study in gastric cancer (orphan condition)"

Authors: Tina van der Horst (submitting author), Michael Budde, Alexander Strasak, Máximo Carreras

A seamless phase II/III study has been set-up by the company F. Hoffmann-La Roche Ltd. to evaluate the efficacy and safety of trastuzumab emtansine (T-DM1) compared to control in patients with HER2-positive advanced gastric cancer. At the start of the trial, patients were randomized to one of three treatment arms: Dose regimen 1; Dose regimen 2; Control. At the end of the first stage of the study in October 2013, the dose and schedule to be used in the second stage of the study were selected by an independent Data Monitoring Committee (iDMC). Since the study's primary endpoint, overall survival (OS) was expected to be immature at time of regimen selection, surrogate information was provided to the iDMC to be used to improve the regimen-selection process. The statistical inference for the selected group at end of the study will use data from both stages, which is possible due to an innovative design concept and statistical testing procedure as discussed in Jenkins et al (2011). The presentation provides an introduction to the study design and methods used. Pros and cons compared to separate phase II and III studies are discussed.

Jenkins M, Stone A, Jennison C. An adaptive seamless phase II/III design for oncology trials with subpopulation selection using correlated survival endpoints. Pharm Stat 2011; 10:347−56

3. Annette Kopp-Schneider (German Cancer Research Centre): "Outcome-adaptive interim monitoring in the INdividualized Therapy FOr Relapsed Malignancies in Childhood Phase I/II (INFORM2) trial series "

Annette Kopp-Schneider(1), Ulrich Abel(2), Ruth Witt(2), Cornelis van Tilburg(2), Olaf Witt(3)

(1)Division of Biostatistics, German Cancer Research Center (DKFZ), Heidelberg

(2)Clinical Trial Center, National Center for Tumor Diseases, Heidelberg

(3) Clinical Cooperation Unit Pediatric Oncology, German Cancer Research Center (DKFZ), Heidelberg

The INFORM program aims at translating next generation molecular diagnostics into a personalized, biomarker driven treatment strategy for children, adolescents and young adults with relapsed or refractory malignant disease of a high-risk entity. The overall concept is divided into 3 steps: pilot phase, registry study and a series of interventional phase I/II trials. The pilot study and the nation-wide INFORM-Registry have established the logistic, clinical, and molecular diagnostic infrastructure for implementing next generation sequencing technology into clinical practice. The next step is a phase I/II biomarker-driven basket trial series (INFORM2) to determine preliminary efficacy of a biomarker driven combination regimen of targeted compounds in children with relapsed malignancies and determine dosing and safety parameters of the combination treatment. Patients will be allocated to a given treatment trial based on presence of a matching molecular alteration.

In this setting, interim analyses will be performed to allow for early stopping, either for futility and/or efficacy. The use of Bayesian posterior and predictive probabilities as decision rule for early stopping has been suggested in the last decade especially for the context of biomarker-targeted therapies with small numbers of patients. We investigate the operating characteristics of a number of stopping criteria for a one-arm trial with dichotomous endpoint. Criteria are based on the Bayesian posterior probability that response probability exceeds a prespecified threshold. In addition, Bayesian predictive probabilities will be evaluated. The final choice of stopping criterion will depend on the principal investigator's preference on the basis of the design's operating characteristics.

Young Statisticians: Practical Application of Bayesian Methods

This session is aimed at statisticians (presenters and attendees) with less than 5 years' experience working in the pharmaceutical industry.

1. Maria Costa (GSK): "Recent Experiences with Implementing Bayesian Designs and Interim Analyses in Early-Phase Drug Development"

In the early stages of drug development, where the focus is in learning rather than confirming, the Bayesian inference paradigm offers intuitive probabilistic statements as an alternative to traditional hypothesis testing. One of the key challenges when designing a clinical trial using Bayesian methodology is the communication with clinical teams who may not be familiar with the concept. This talk will cover some recent experiences with implementing Bayesian decision rules in early phase clinical trials. In particular, how clinical teams were introduced to the concept of interim analysis using Bayesian predictive probabilities as a risk mitigation strategy, and which graphical tools proved useful to support decision-making.

2. Daniel Sabanés Bové (Roche): "Bayesian decision criteria with two efficacy endpoints in early phase trials"

Decision making for early phase trials usually is based on a single efficacy endpoint, for example objective response rate (ORR) in oncology trials. There is now an extensive literature on monitoring ORR with Bayesian inference, see e.g. Saville et al (2014, Clin Trials), in order to facilitate efficient and good decisions.

However, for drug combinations or for cancer immunotherapies, the clinical benefit for the patients may not be captured completely by ORR. Therefore, long-lasting stabilization of the disease (incorporated in the disease control rate, DCR) or extended duration of response (DoR) are therefore considered and taken into account informally for decision making. The rationale is that durable stable disease can represent anti-tumor activity (see e.g. Wolchok et al, 2009, Clin Can Res) and extended DoR capturing deeper responses could translate to longer overall survival.

Here we present a Bayesian decision making framework for two "co-primary" efficacy endpoints in early phase trials, and give details on the joint ORR/DCR and ORR/DoR monitoring. The efficacy can either be compared to fixed thresholds in the two dimensional

outcome space, or be compared to data from relevant competitor drugs. Posterior and predictive probabilities can be utilized for early stopping of the trial. The methodology will be illustrated with an example.

3. Wilmar Igl (AstraZeneca): "Application of Bayesian hierarchical generalized linear models using weakly informative prior distributions to identify rare genetic variant effects on blood pressure"

Background. Currently rare genetic variants are discussed as a source of "missing heritability" of complex traits. Bayesian hierarchical models were proposed as an efficient method for the estimation and aggregation of conditional effects of rare variants. Here, such models are applied to identify rare variant effects on blood pressure.

Methods. Empirical data provided by the Genetic Analysis Workshop 19 (2014) included 1,851 Mexican-American individuals with information on diastolic blood pressure (DBP), systolic blood pressure (SBP), and presence of hypertension (HTN) and 461,868 variants from whole-exome sequencing of odd-numbered chromosomes. Bayesian hierarchical generalized linear models using weakly informative prior distributions were applied.

Results. Associations of rare variants chr1:204214013 (estimate = 39.6, Credible Interval (CrI) 95% = [25.3, 53.9], Bayesian p = 6.8 × 10−8) in the PLEKHA6 gene and chr11:118518698 (estimate = 32.2, CrI95% = [20.6, 43.9], Bayesian p = 7.0 × 10−8) in the PHLEDB1 gene with DBP were identified. Joint effects of grouped rare variants on DBP in 23 genes (Bayesian p = [8.8 × 10−14, 9.3 × 10−8]) and on SBP in 21 genes (Bayesian p = [8.6 × 10−12, 7.8 × 10−8]) in pathways related to hemostasis, sodium/calcium transport, ciliary activity, and others were found. No association with hypertension was detected.

Conclusions. Bayesian hierarchical generalized linear models with weakly informative priors can successfully be applied in exome-wide genetic association analyses of rare variants.

Note: The presentation is based on a master thesis in medical statistics which was awarded with the PSI University Prize 2015 (University of Leicester).

Poster Session

Join us for drinks, canapés and the opportunity to speak to the authors of the selection of posters on display. This event is kindly sponsored by Roche.

Tuesday:

Methodological challenges in the (added) benefit assessment of drugs

Opening the second day of the conference is Prof. Dr. med. Stefan Lange, Deputy Director at the Institute for Quality and Efficiency in Health Care (IQWiG), the German agency responsible for assessing the quality and efficiency of medical treatments, including drugs, non-drug interventions, diagnostic and screening methods, and treatment and disease management.

Since 2011, new drugs entering the German market must undergo an early benefit assessment. Within this assessment the new drug is compared with the generally accepted and best-tested (evidence-based) treatment standard in the respective therapeutic indication. To conclude an added benefit of the new drug, patient-relevant advantages must be shown for mortality, morbidity or quality of life. In addition, the extent of added benefit must be quantified and the likelihood ("probability") of added benefit determined. In the upcoming presentation, methods developed by the Institute for Quality and Efficiency in Health Care (IQWiG) to determine the extent of added benefit [1] will be presented and discussed within the context of other current proposals [2,3]. Furthermore, challenges will be addressed concerning the consideration of surrogate outcomes, indirect comparisons, the specific problem of treatment switching in oncology trials, and the analysis of adverse events.

[1] Skipka G, Wieseler B, Kaiser T, Thomas S, Bender R, Windeler J, Lange S. A methodo-logical approach to determine minor, considerable and major treatment effects in the early benefit assessment of new drugs. Biom J. 2016; 58: 43-58.

[2] Cherny NI, Sullivan R, Dafni U, Kerst JM, Sobrero A, Zielinski C, et al. A standardised, generic, validated approach to stratify the magnitude of clinical benefit that can be anticipated from anti-cancer therapies: the European Society for Medical Oncology Magnitude of Clinical Benefit Scale (ESMO-MCBS). Ann Oncol. 2015; 26: 1547-73.

[3] Schnipper LE, Davidson NE, Wollins DS, Tyne C, Blayney DW, Blum D, et al. American Society of Clinical Oncology Statement: A Conceptual Framework to Assess the Value of Cancer Treatment Options. J Clin Oncol. 2015; 33: 2563-77

Benefit Risk Assessment within Health Technology Assessment: Experiences and Opportunities

1. Friedhelm Leverkus (Pfizer): "Design of Phase III Studies to get an additional Benefit in AMNOG"

In 2010, the Federal Parliament (Bundestag) of Germany passed a new law (Arzneimittelmarktneuord- nungsgesetz, AMNOG) on the regulation of medicinal products that applies to all pharmaceutical products with active ingredients that are launched beginning January 1, 2011. The law describes the process to determine the price at which an approved new product will be reimbursed by the statutory health insurance system. The process consists of two phases. The first phase assesses the additional benefit of the new product versus an appropriate comparator (zweckma¨ßige Vergleichstherapie, zVT). The second phase involves price negotiation. Focusing on the first phase, we will investigate the requirements of benefit assessment of a new product under this law with special attention on the methods applied by the German authorities on issues such as the choice of the comparator, patient relevant endpoints, subgroup analyses, extent of benefit, determination of net benefit, primary and secondary endpoints, and uncertainty of the additional benefit. These requirements may have an influence on the design of the clinical studies.

2. Susan Talbot (Amgen): "Benefit risk assessment within health technology assessment"

Although focusing on clinical aspects, benefit-risk assessment (BRA) of a healthcare product can directly inform health technology assessment (HTA). HTA is multidisciplinary; and examines broader aspects of a health technology including safety, effectiveness, cost-effectiveness and other social aspects.

HTA agencies have already been applying qualitative and quantitative BRA approaches during their evaluation of new drug submissions for pricing and reimbursement approvals. However, to date, there are no standard requirements for review and approval processes. Similarly, industry-agreed standards are also lacking. Recent research from various initiatives has made significant progress in BRA and HTA, including development of structured and quantitative approaches, visual tools, decision support tools, and multi-criteria decision analysis (MCDA) methods.

In this session, we will present an up-to-date review of the outputs from key initiatives in this area, including those from the EFSPI joint working group of BRA and HTA SIGs, and then discuss in more detail HTA and BRA approaches such as the QALY and incremental net benefit based on the NICE approach, the ASCO value framework, and other pragmatic approaches such as benefit-risk ratio/difference and clinical utility indices and links between them. Finally, we will systematically compare the approach taken to BRA between HTA authorities and regulatory agencies, and comment on the challenges to, and impact of the statisticians in the pharmaceutical industry in these emerging but converging fields.

3. Fabian Volz: "Could Multiple-Criteria Decision Analysis be used in the German Benefit Assessment?"

Since 2011 every new pharmaceutical in Germany has to undergo an early benefit assessment to demonstrate superiority versus an appropriate comparator in patient-relevant endpoints (additional benefit) as a door opener and a basis for price negotiations with the Statutory Health Insurance Funds. The Institute for Quality and Efficiency in Health Care (IQWiG) conducts this assessment based on the submitted benefit dossier by the pharmaceutical company and the Federal Joint Committee (G-BA) decides afterwards about the extent and probability of the additional benefit. This additional benefit usually consists of positive and negative effects in single considered endpoints. Lacking a clear methodological approach these endpoints are balanced and combined in a semi-quantitative manner by IQWiG and G-BA in one overall additional benefit. For this balancing methods from the multiple-criteria decision analysis (MCDA) could be used to make this process more transparent and comprehensible for all involved parties.

Translational Biomarkers: from Preclinical to Phase 1

The National Institutes of Health (NIH) defines translational research as the movement of discoveries in basic research to application at the clinical level. In this session we will see examples of how pre-clinical data has been optimally utilized to enhance phase I trials. Please see the abstracts for more details and we hope to see you at this session.

Speakers:

1. Alun Bedding (AstraZeneca): "Translation of Pre-Clinical Pharmacokinetic and Pharmacodynamic Parameters in the Determination of Dosing and Dose Response in First Time in Human Studies"

Decisions based on human pharmacokinetic predictions have a significant impact on the candidate selection of a drug, as well as on the number of animal studies triggered to support it. The prediction of the human PK data is therefore key, but how well is it done?

There is much debate within the pharmaceutical industry about which method to predict human PK parameters (e.g. clearance, half-life, volume of distribution, bioavailability, dose...) from in vitro or in vivo animal data should be used, and the reliability of those predictions. The accuracy of such predictions is important given the doses for first time in human studies are based on them. Recently the PhRMA CPCDC Initiative on Predictive Models of Human PK published its recommendations on this issue.

In addition, it has long been argued that animal data does not translate well to humans. In some therapeutic areas it is possible the animal models of dose response are similar to those of humans and therefore it would make sense to use them.

This talk will focus on how pre-clinical data can be used to maximise the information in a first time in human study to make for efficient dose escalation.

2. Claire Brittain (Eli Lilly): "Translation, a two way street - A case study of Event Related Potentials (ERPs) as a neuroscience biomarker"

In early clinical drug development, biomarkers capable of providing proof of mechanism are considered critical tools and can help reduce attrition during phase II clinical trials. However, with neuroscience drugs it's common to use different measures in rats and humans e.g. Water mazes in rats and ADAS-Cog questionnaires for Alzheimer's in humans. They are both excellent measures in their own right but translation can be greatly improved if you start by comparing apples with apples.

This gives us 3 options:

1. Ask volunteers to swim in circular tanks looking for a hidden platform

2. Teach rats to answer complex questions on their cognitive impairment

3. Find a new measure that can used in both species and thus allows us to compare directly

This presentation will take you through the journey of how we selected our biomarker (Auditory Evoked Related Potentials) and the considerations in designing and analysing the experiments between species.
I hope to show what can be achieved if we blur the hard line between non-clinical and clinical and think of it more of an iterative discussion.

3. Thomas Jaki (Uni. of Lancaster) : "Improving Design, Evaluation and Analysis of Early Drug Development Studies (IDEAS)"

Drug development is a long and costly process which suffers from the major shortcoming that frequently failure is often only determined during the final stage. Advanced statistical methods for study design, evaluation and analysis, employed already at the early phases of drug development, have a great potential to increase the efficiency of the development process.

IDEAS is a European training network for 14 early stage researchers working on statistical methods for early drug development. The network is funded by the European Union's Horizon 2020 research and innovation programme and comprises of 8 full partners and three associated partners at major European universities, the pharmaceutical industry, and consulting companies.

In this talk we will outline the structure of IDEAS and highlight two specific projects that are focusing on translation between pre-clinical and clinical studies.

Statistical Challenges Relating to Safety

1. Elizabeth Merrall (GSK): "Signal detection experiences using clinical trial data"

Clinical Data Repository (CDR) is the GSK Vaccines (formerly Novartis Vaccines) integrated environment for storing, managing and reporting clinical trial data (and metadata) based on SAS® Drug Development (SDD) version 3.5. CDR has been developed to revolutionize our ability to:

- *Review all available safety data, via merging, in real-time*
- *Address complex health authority questions quickly and completely*
- *Produce CDISC-compliant submissions*
- *Mine our overall database for scientific and commercial queries*
- *Improve overall productivity in Clinical Research & Development*

Through CDR, clinical trial data are now readily available in CDISC SDTM format for more than 230 of our studies. This resource provides excellent insight into our products and an opportunity to improve pharmacovigilance (PV), namely detecting safety signals, much earlier than usual in the drug/vaccine development process.

In this paper, we present our experiences of applying frequentist and Bayesian statistical signal detection approaches to clinical trial data. These experiences stem from extensive collaboration with PV physicians and the keen interest in monitoring potential safety issues, and now thanks to CDR, also the possibility to carry out periodic reviews to detect unknown issues.

2. Arthur Allignol (Universität Ulm): "Statistical issues in the analysis of adverse events in time-to-event data"

The aim of this work is to shed some light on common issues in the statistical analysis of adverse events (AEs) in clinical trials, when the main outcome is a time-to-event endpoint. To begin, we show that AEs are always subject to competing risks. That is, the occurrence of a certain AE may be precluded by occurrence of the main time-to-event outcome or by occurrence of another (fatal) AE. This has raised concerns on "informative" censoring. We show that neither simple proportions nor Kaplan-Meier estimates of AE occurrence should be used, but common survival techniques for hazards that censor the competing event are still valid, but incomplete analyses. They must be complemented by an analogous analysis of the competing event for inference on the cumulative AE probability. The commonly used incidence density (or: incidence rate) is a valid estimator of the AE hazard assuming it to be time-constant. An estimator of the cumulative AE probability can be derived, if the incidence density of AE is combined with an estimator of the competing hazard. We discuss less restrictive analyses using non- and semi-parametric approaches. We first consider time-to-first-AE analyses and then briefly discuss how they can be extended to the analysis of recurrent AEs. We will give a practical presentation with illustration of the methods by a simple example.

3. Katie Patel (Roche): "Could Exposure/QT Response Analysis in Early Clinical Studies Replace the Thorough QT Study?"

The IQ-CSRC prospective study provides evidence that robust QT assessment in early-phase clinical studies can replace the thorough QT study (1).

We retrospectively applied the approach proposed by the IQ-CSRC, using ECG and drug concentration data collected during the single ascending dose phase of an early clinical study. We present the results of our analysis and highlight points for consideration when planning future such analyses. Based on our results, we recommend that project teams consider including ECG monitoring during early phase studies and pre-specify an exposure response analysis.

(1): Darpo, B. et al. Results From the IQ-CSRC Prospective Study Support Replacement of the Thorough QT Study by QT Assessment in the Early Clinical Phase. Clinical Pharmacology and Therapeutics, 97:4, 326-335 (2015).

Using Bayesian Methods to Predict Success

1. Phil Woodward (Pfizer): "Bayesian Methods and Thinking in Pfizer's Early R&D Division."

Over the last seven years Pfizer have greatly increased the use of Bayesian methods and thinking in a major Division of their early Research and Development organisation. We now have formalised goals to utilise Bayesian methods in all Proof-of-Mechanism and Proof-of-Concept studies unless there is good reason not to do so. This presentation explains how the change came about, discussing the motivation for change and the management leadership required. We share our experience in preparing both the statisticians and our colleagues for the new approaches, highlighting the major hurdles that have been overcome as well as those that still exist. Some examples of the benefits already realised are shared, as well as a discussion of what the future might hold for even greater use of Bayesian methods

.

2.Christian Röver (University Medical Center Göttingen): "Bayesian random-effects meta-analysis made simple"

During the clinical development of a new drug it is often necessary to merge information from few trials in order to assess the available evidence [1]. Similarly, in the context of rare diseases, evidence commonly is based on the meta-analysis of only a handful of trials. In such cases, Bayesian methods allow to coherently infer the joint outcome while accounting for potential heterogeneity between studies, as the small number of studies does not pose a problem, and additional external evidence may be accounted for [2,3]. Using some examples, here we demonstrate how random-effects meta-analysis may easily be performed using a new R package, without having to worry about computational issues. The use of MCMC is avoided by use of numerical integration techniques implemented in the R package resulting in an easy to use tool for Bayesian random-effects meta-analysis.

[1] European Medicines Agency (EMEA) (2001). Points to consider on application with 1. meta-analyses; 2. one pivotal study. CPMP/EWP/2330/99.

[2] Smith T.C., Spiegelhalter D.J., Thomas A. Bayesian approaches to random-effects meta-analysis: A comparative study. Statistics in Medicine 1995;14(24):2685–99.

[3] Friede, T., Röver, C., Wandel, S., Neuenschwander, B.: Meta-analysis of few small studies in orphan diseases (2015). (submitted)

3. Kaspar Rufibach (Roche): "Bayesian Predictive Power: Choice of Prior and some Recommendations for its Use as Probability of Success in Drug Development"

Bayesian predictive power, the expectation of the power function with respect to an assumed distribution on the true underlying effect size, is routinely used in drug development to quantify the probability of success of a clinical trial. Choosing the prior is crucial for the properties and interpretability of Bayesian predictive power. We review recommendations on the choice of prior for Bayesian predictive power and explore its features as a function of the distribution on the true underlying effect. The density of power values induced by a given prior is derived analytically and its shape characterized. This characterization is used to show that summarizing the power value density in one number, the mean, might not fully capture the features of that distribution. Conditions under which this summary statistic is more sensible for a Normal prior are derived. Alternative priors are proposed and practical recommendations to assess the sensitivity of Bayesian predictive power to its input parameters are provided.

Creative Designs: Time for Something a Little Bit Different?

We are always being asked to be more innovative in our jobs as statisticians. This session aims to share three different examples of people thinking differently with the problems of clinical trials and analysis. Please read the abstracts to find out more and we hope to see you in this session.

1. Aaron Dane (DaneStat Consulting): "The Application Platform Trials to Antibiotic Development"

Platform trials allow the study of multiple treatments. Recruitment continues beyond the evaluation of a single treatment with therapies entering the trial when available, and leaving the trial when their evaluation is completed. Data from all patients are incorporated into a statistical model that can be used to define future treatment allocation and provide likelihood based decisions for when each treatment should leave the trial having shown sufficient efficacy or a lack of efficacy, making further study unnecessary (futility).

The benefits of performing such trials are that they allow more efficient use of resources by creating an effective clinical trial network, sharing a common control to assess each new antibiotic, and through the application of Bayesian hierarchical modelling to "borrow" information across body sites.

Such trials are being considered as a way to provide confirmatory regulatory evidence for antibiotic development for multi-drug resistant pathogens where it is challenging to recruit a large number of patients. These trials can be used to study multiple antibiotics across multiple infection sites, and although they can help make antibiotic development more feasible, a number of assumptions are necessary to allow information to be combined.

This talk will outline how these assumptions have been considered, whether the degree of uncertainty associated with any estimate is appropriate when providing confirmatory evidence, and will outline how this approach is being piloted in a cross-industry / academia antibiotic trial.

2. Jürgen Hummel (PPD): "SPCD: a study design option for high placebo response"

Placebo response is high and rising in clinical trials in many indications, impairing the detection of an accurate therapeutic signal. Subsequently a high proportion of studies in Phase II and III are negative or inconclusive. Various options have been tried over the years to address that, including designs using a placebo lead in or randomised withdrawal. The presentation introduces another option, the Sequential Parallel Comparison Design (SPCD/Trimentum™), a two-stage design and analysis method that combines the Stage 1 data from all-comer subjects with the Stage 2 data from placebo non-responders identified in Stage 1. Different design variants are shown and a variety of analysis options are discussed, as well as extensions of the SPCD.

3. Sebastian Straube (University of Alberta): "Enriched enrolment randomised withdrawal trial designs (EERW)"

Enriched enrolment trial designs have been utilized for over 40 years and have recently come into the focus of attention. An area of medicine where such trial designs have been increasingly employed in the recent past is the treatment of chronic painful conditions. Here we will make reference to pain trials to outline the principles of enriched enrolment designs of randomized controlled trials. Partial enriched enrolment and complete enriched enrolment will be defined. We will examine the practice, reporting and consequences of different enrolment strategies and also discuss criteria for risk of bias assessments of trials using such enriched enrolment strategies. Recommendations will be made for the conduct and reporting of enriched enrolment trials.

Recurrent Event Analysis

In many chronic disease trials, treatment effects have been assessed through time-to-event endpoints; e.g. time to death or hospitalization in chronic heart failure trials or time to tumour occurrence in cancer trials. Traditionally, these time-to-event endpoints have been analyzed using a time-to-first-event analysis approach. A recurrent event approach, in which the first as well as subsequent events are included, would more accurately reflect the true burden of the illness on the patient. In this session we will discuss the opportunities and challenges in the use of recurrent event endpoints and touch on design aspects.

Speakers:

1. Jennifer Rogers (University of Oxford): "Sample Size Considerations for the Analysis of Recurrent Events"

Many chronic diseases are characterised by nonfatal recurrent events. Examples of such include asthma attacks in asthma, epileptic seizures in epilepsy and hospitalisations for worsening condition in heart failure. Analysing all of these repeat events within individuals is more representative of disease progression and more accurately estimates the effect of treatment on the true burden of disease. Analyses to date of recurrent heart failure hospitalisations have typically been retrospective and post-hoc, and not pre-specified primary analyses. These analyses of recurrent heart failure hospitalisations may challenge current guideline interpretations, but guideline writing committees must give most credence to the pre-specified primary analysis of trials and be cautious about retrospective analyses.

Previous analyses of recurrent heart failure hospitalisations have advocated the use of the negative binomial distribution, but sample size calculations for the negative binomial model are not well documented. This is likely to be a major limiting factor in getting recurrent events considered as a pre-specified primary outcome in the design of future clinical trials of heart failure. This talk shall demonstrate the use of the negative binomial sample size formula in practice and shall discuss the associated assumptions.

2. Arno Fritsch (Bayer) & Patrick Patrick Schlömer (Bayer): "Recurrent events vs. time-to-first event in heart failure trials – Simulations investigating power and sample size"

In the past, the standard (composite) primary endpoint in heart failure (HF) trials was the time to either HF hospitalization or cardiovascular death, whichever occurs first. With improved treatments on the market HF has now changed from a short-term and quickly fatal condition to a chronic disease, characterized by recurrent HF hospitalizations and high mortality. Therefore, there is interest to move from the standard 'time-to-first' event to recurrent events as the primary efficacy endpoint. Through this on hopes to better characterize and quantify the full disease burden of HF because recurrent hospitalizations have a substantial impact on both the patients' well-being and the health systems. Moreover, one expects to achieve practical gains by means of sample size savings due to incorporating more statistical information.

As the literature on power and sample size calculations for recurrent event analyses is limited and does e.g. not capture the specific feature of combining recurrent HF hospitalizations and the terminal event cardiovascular death to one endpoint, we conducted an extensive simulation study to quantify potential power gains and sample sizes

savings by using recurrent events instead of standard 'time-to-first' event methods. We investigate the influence of different factors such as event rates, treatment discontinuation and inter-patient heterogeneity. Finally, we examine the advantages and disadvantages of using recurrent event methods, also with respect to practical implications.

3. Philip Hougaard (Lundbeck): "Statistical analysis of recurrent events, based on simple frailty models, and extensions"

Recurrent events refer to multiple occurrences of some event, for example, epileptic seizures, hypoglycemic episodes in diabetes or heart attacks. The actual data may refer to a count of events during an interval, or to data of times when the event occurs. Studying only the time to the first event is possible by means of standard survival data methods, which is quite simple but in many cases also unsatisfactory. Studying all occurrences gives a more complete picture of the disease burden. At the same time, it gives a more precise picture but at the price of using more complex statistical methods that adequately account for the relevant sources of random variation. Typically, there is subject variation, that is, the event rate differs between subjects, for example, due to unobserved risk factors. Such overdispersion may be modelled by a frailty model, where the frailty is a random term describing the effect of individual unobserved risk factors and the events then occur according to a Poisson process conditional on the frailty. The simplest case is when the frailty is constant and follows a gamma distribution, in which case, the number of events in an interval follows a negative binomial distribution. Obvious extensions of this model are obtained by using other distributions than the gamma and allowing for observation periods that differ between individuals. I will focus on a third type of extension, namely releasing the assumption of the frailty being constant. This extension is relevant for a cross-over type experiment as well as for a deeper study of the dependence over time that could apply for a clinical study with several phases, such as titration and maintenance.

Personalised Medicine Tutorial

Sandeep Menon, Vice President and Head of Statistical Research and Consulting Center, Pfizer will be running a 90 minute tutorial about the concepts and statistical methodology related to Personalized Medicine

Title: Overview of Statistical and Design Considerations in Personalized Medicine

Abstract: Personalized medicine is described as providing "the right patient with the right drug at the right dose at the right time for the right outcome." Personalized medicine is a relatively young but rapidly evolving field of clinical research. It involves identifying genetic, genomic, and clinical characteristics that have the potential to accurately predict patient's susceptibility of developing a certain disease and its response to treatment. Personalized medicine is the translation of this knowledge to patient care. However, this "translation" can be very challenging in the phase of limited knowledge of the biomarker and /or appropriate diagnostics. Hence, the appropriate selection of the study design is important to critically determine biomarker performance, reliability and eventually regulatory acceptance. This introductory 90 minute tutorial will provide a general overview of the concept and statistical methodology related to personalized medicine. Specifically, it will discuss various designs including adaptive designs available at our disposal and its merits and limitations. Case studies will also be discussed.

Optimizing Drug Development when Patients are Hard to Find

This session takes both a methodological and a practical look at how to conduct studies in small populations / rare diseases. Speakers will explore the challenges from both these perspectives and suggest ways forward to improve the efficiency of such trials.

Speakers:

1.Thomas Jaki (University of Lancaster): "Clinical Trial Design for Rare Diseases using Bayesian Bandit Models"

Development of treatments for rare diseases is particularly challenging. Learning about treatment effectiveness with a view to treat patients in the a population outside the trial, as in the traditional fixed randomised design, is less important as often a large part of the whole patient population is participating in the trial. Now, the priority is to treat the patients within the trial as effectively as possible whilst still identifying the superior treatment. This problem is a natural application area for bandit models which seek to balance the underlying exploration versus exploitation trade-off inherent in clinical trial design.

We formulate this model as a finite-horizon Markov decision problem and use dynamic programming (DP) to obtain an optimal adaptive treatment allocation sequence which maximises the total expected reward over the planning horizon. However, this optimal design is deterministic, which is undesirable from a practical point of view, so we modify it by forcing actions to be randomised. Further concerns with this design are that it is underpowered and there is the possibility that all patients will be allocated to only one of the treatments. To resolve these issues, we propose a constrained version of the randomised DP design. We evaluate several performance measures of these designs through extensive simulation studies. For simplicity, we consider a two-armed trial with binary endpoints and assume immediate responses.

Simulation results for this constrained variant show that (i) the percentage of patients allocated to the superior arm is much higher than in the traditional fixed randomised design and (ii) relative to the optimal DP design, the power is largely improved upon. Furthermore, this design exhibits only a small bias of the treatment effect estimator, and has the desirable property that changing the degree of randomisation does not impact the results significantly.

2. Nikos Sfikas (Novartis): "Challenges in developing medications for small populations/rare diseases – an industry statistician's perspective"

Feasibility of conducting well-controlled and sufficiently powered studies in small populations/rare diseases is the biggest challenge that a company will face when pursuing development of a medication in such populations/diseases.

Substantial number of initiatives are currently ongoing that aim to provide the methodology for more efficient designs, that will help to facilitate conducting better studies in the small populations/rare diseases setting.

Still, a statistician faces several additional challenges, when supporting clinical teams to design such studies. Negotiations within the company as well as with external stakeholders on study design characteristics, endpoints definition, analysis approaches for such trials, level of evidence required to support decision making, are areas where a statistician should give particular emphasis on. This talk discusses some of the challenges a biostatistician working for a pharmaceutical company should expect to face when developing medications for rare diseases/small populations.

3. Thomas Zwingers (CROS NT): "The use of historical controls in randomized clinical trials in rare diseases"

The use of prospectively randomized clinical trials is the gold standard to show efficacy in clinical research when evaluating new compounds. Especially in rare diseases, e.g. in cancer subtypes like brain

tumours, it is sometimes very difficult to recruit the requested number of patients in a reasonable timeframe. On the other hand, there are often registers on such patients which have been treated with the standard of care or individual treatment regimens.

Stuart Pocock [1] suggested a compromise in the way to combine historical controls with a prospective control group in a randomized clinical trial.

We implemented this design in a clinical trial in patients with a rare tumour in order to overcome the problems recruitment and limited resources.

We will discuss the problems which are inherent to historical data not collected prospectively according to a study protocol and the assumptions of the sample size calculation.

We will show that the inclusion of historical controls substantially reduced the number of patients to be recruited and the total duration of study.

[1] Stuart J. Pocock: The combination of randomized and historical controls in clinical trials. J Chron Dis 1976, Vol.29

Breaking New (Statistical) Ground – RSS/PSI Prize Winner 2015

The PSI/RSS Award for Statistical Excellence in the Pharmaceutical Industry recognises an outstanding level of influence in the application of an existing statistical practice, or an innovation, that has strengthened the quality and efficiency of investigations in the pharmaceutical industry. For the first time in its five-year history, the Award was given to two joint winners, both of whom are joining us in Berlin to present their work. PSI would like to congratulate both winners.

1. Mark Whitlock (Pfizer): "Driving the Robustness of Preclinical Research within the Pharmaceutical Industry"

It is unusual to pick up a recent issue of Nature or Science that doesn't include an article on the issue of non-reproducible research. It would seem that research is plagued by findings that are not reliable and cannot be reproduced. The pharmaceutical industry is not immune to these same issues. Replication of published research findings is a key component of drug target identification and provides confidence to progress internal drug projects. Additionally, we use data from internal assays to assess the biological and pharmacokinetic activity, selectivity and safety of novel compounds, and to make decisions that impact their progression towards clinical development. While pharmaceutical companies often employ statisticians specifically to engage with research scientists, the ratio of statisticians to scientists is typically low. This talk will describe the role of a preclinical statistician, outline the key challenges they face and focus on the Assay Capability Tool. The ACT was created by Research Statistics within Pfizer to guide the development of drug discovery assays and to address issues of robustness and reproducibility in research. It promotes easy to follow but absolutely essential experimental design strategies and represents the distilled experience of the provision of over three decades of statistical support to laboratory scientists.

2. Nicky Best (GSK): "Using prior elicitation and Bayesian thinking to help shape decision making in drug development"

Since 2014, clinical statisticians at GlaxoSmithKline (GSK) have been using prior elicitation techniques to enable quantification of existing knowledge in the absence of directly relevant data, and to help predict probability of success of next study(s) at key milestone decision points for all phases of clinical drug development. This initiative forms a key component of an R&D-wide focus on innovation in clinical design at GSK, which aims to establish Bayesian approaches and use of prior distributions as standard practice to support internal decision-making and analysis.

In this talk, I will give a flavour of what the prior elicitation process involves and how the elicited priors have been used at GSK, e.g. to quantitatively choose between competing clinical trial designs for the next stage of drug development, to explore staged development activities and to determine the merits of interim/futility assessments. I will also discuss some of the challenges along the way, including first having to teach ourselves on both the statistical and psychological aspects of an elicitation process, then educating our clinical colleagues and senior management about this, as well as encouraging and influencing project teams to work with us and to use our material in their decision-making and investment review meetings.

Subgroups – Divide and Conquer?

The scientific best practices and regulatory views on the use (and abuse) of subgroups have undergone considerable recent evolution.

This interactive session will include presentations on ICH E17 (General principle on planning/designing Multi-Regional Clinical Trials) and also EMA's recent subgroup guideline. Our three highly experienced speakers will also be leading a lively panel discussion towards the end of the session, where the implications of this new guidance can be dissected.

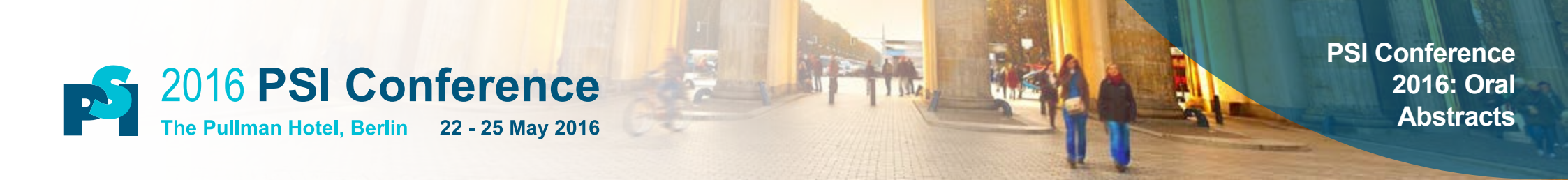
Speakers:

1. Armin Koch (Medizinische Hochschule Hannover): "Regulatory Considerations for Regional and Subgroup Analyses"

In the recent past two regulatory documents have been announced (or provided as a draft) that are supposed to clarify the importance of subgroups of randomized clinical trials for drug licensing. These are the "Guideline on the investigation of subgroups in confirmatory clinical trials" published by the EMA and "General Principles for Planning and Design of Multi-Regional Clinical Trials (ICH-E17)" developed in the framework of the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. Whereas the latter is the first document to discuss the role of subgroups in the review of clinical trials and their importance in benefit/risk assessment, the latter provides argumentation, under which circumstances subgroups of a multi-regional clinical trial may suffice as a basis for regulatory decision making in regions of the world. Obviously both topics are interrelated and this presentation is meant to provide a frame for the discussions in this session.

2. Aaron Dane (DaneStat Consulting): "Subgroup Analysis Approaches to Inform Regulatory Labelling"

EMA has issued a draft guidance on approaches to subgroup analysis, and convened a workshop to discuss this guidance in 2014. As this guidance provides a very useful framework for interpreting subgroup effects, but does not address the question of how to assess consistency and when results are likely to affect regulatory labelling, a PSI Working Group has been formed. This WG has evaluated various statistical approaches to assess subgroup effects, and have discussed their implications for regulatory labelling with some EMA statistical regulators. This talk will summarise the approaches evaluated and the progress made in terms of their possible regulatory acceptance.



3. Byron Jones (Novartis): "Multiregional Trials from the Perspective of a Subgroup Analysis"

The regions in a multiregional trial are, by definition, predefined and obvious. This means that the usual problem of discovering unknown subgroups in a large dataset is absent in this case. What is an issue, however, is the interpretation of significant differences between subgroups (regions). A related issue is how regulatory agencies view the results of a multiregional trial with reference to their own particular region. We will review the issues involved in the planning and analysis of multiregional trials and give examples of where a significant treatment-by-region interaction has been a major cause of concern.

Latest Updates on Estimands

Defining the primary objective of a clinical trial in the presence of non-compliance or non-adherence to the assigned treatment is crucial for the choice of design, the statistical analysis and the interpretation of the results. This raises the need for a structured framework to specify the primary estimand (i.e. "what is to be estimated"). The missing data report released in 2010 by the National Academy of Science, "Prevention and Treatment of Missing Data in Clinical Trials", recommends explicit specification of a casual estimand in the protocol of a confirmatory trial. This is also reflected by the decision of the International Conference on Harmonization (ICH) to amend its E9 guidance in the coming years to discuss estimands and their role in clinical trials. The focus of this session is to discuss the draft ICH E9 addendum (which is expected to be released end of 2016) from an industry and regulatory viewpoint. The estimand framework will be illustrated based on case studies.

Speakers:

1.Frank Pétavy (EMA): "Latest Update on ICH E9(R1) Addendum Development and Views from EU Regulators"

Defining the primary objective of a clinical trial in the presence of non-compliance or non-adherence to the assigned treatment is crucial for the choice of design, the statistical analysis and the interpretation of the results. It is thought that many clinical trial protocols fail to give a precise definition of what is to be estimated, the estimand. However, there is no definitive guidance available on what constitutes an appropriate primary estimand for a confirmatory clinical trial. As a result the ICH Steering Committee endorsed a final Concept Paper in October 2014 with the goal of developing a new regulatory guidance, suggested to be an Addendum to ICH E9. The aim of the addendum is to promote harmonised standards on the choice of estimands in clinical trials and an agreed framework for planning, conducting and interpreting sensitivity analyses of clinical trial data. A working group sponsored by ICH was established in 2014 to develop the addendum. This talk will provide an update of activities of the ICH E9(R1) expert group and provisional plans for publication of the draft addendum. Feedback will be given from a recent meeting on estimands with clinicians and statisticians from the European medicines regulatory network.

2. Kaspar Rufibach (Roche): "Time to think about estimands for time-to-event endpoints? A few examples"

The current discussion on estimands primarily grew out of the wish to get more clarity on assumptions made when analyzing longitudinal or clustered data with missing values. This will likely also lead to a repositioning of the role of sensitivity analyses. Using case examples from recent Phase 3 clinical trials, we will illustrate that for time-to-event endpoints such as progression-free survival, often the estimand is not clearly specified either and thus results may be subject to ambiguity and not answer the question that scientist were asking.

In one of these case studies, we discuss that Health Authority requests for additional complicated sensitivity analyses during a filing, with potential to delay approval, are often not clearly focused on getting more information on a clearly defined quantity but rather simply vary some assumptions to pressure-test robustness of results. We advocate embedding such questions in the estimand framework in order to get a clear understanding of the value of such additional analyses and why they are needed.

3. Oliver Keene (GSK): "Estimands in confirmatory trials: time to face De Facto"

Recent publications and presentations from statisticians working in regulatory agencies have indicated interest in "de facto" estimands for the primary analysis for confirmatory trials of efficacy. In many trials, the implementation of de facto estimands requires modelling of the missing data. This can be achieved by use of reference-based imputation or by making specific assumptions regarding the outcomes for subjects with missing data. This latter approach leads to tipping point sensitivity analysis designed to address the robustness of analysis to assumptions regarding missing data. These concepts will be illustrated with an example of a trial with recurrent event data.

Case Studies Involving Dose Exposure Modelling

An important goal for clinical studies should be to determine the dose response relationships for both efficacy and safety endpoints. Only with precisely quantified relationships can drug companies and regulators determine a suitable range of doses to consider for approval. This session shares three examples of dose exposure/dose response work.

Speakers:

1. Daniel Sabanés Bové (Roche): "Bayesian decision criteria with two efficacy endpoints in early phase trials"

Model-based dose escalation designs have gained increasing interest due to the need for more efficient and informative Phase I trials. The wide-spread implementation of such designs has been hindered by the need for either licensing specialized commercial software (e.g. FACTS) or programming the design and simulations from scratch for each project. Therefore we developed the R-package "crmPack", which is now publicly available on CRAN. By providing a simple and unified object-oriented structure for model-based dose escalation designs, crmPack enables the use of standard designs, full flexibility to adapt designs to the project, and easy extension to new designs. The framework for the designs comprises the data structure, regression model and prior distribution specification, maximum increment rules, derivation of the next best dose, stopping rules for adaptive sample size, adaptive cohort sizes, and starting dose and dose grid specification. In addition to the classic continual reassessment method (CRM) and escalation with overdose control (EWOC) designs with possibly advanced prior specifications with e.g. mixtures and minimal informative priors, crmPack currently features dual-endpoint (safety and biomarker) designs and two-part (SAD followed by MAD part) designs. As crmPack is actively being developed further, this list will be extended.

We introduce crmPack by outlining the design framework and show how easy it is to specify a design, run simulations, and report dose recommendations during the trial.

2. Alexandra Jauhiainen (AstraZeneca): "Dose-finding for disease with rare event endpoints"

Patients that suffer from asthma and COPD (Chronic Obstructive Pulmonary Disease) experience episodes of severe symptom worsening called exacerbations. These events are one of the main clinical endpoints in both diseases. The problem with designing trials using exacerbations as the endpoint is that the events are rare and occur randomly in time, demanding large and lengthy trials.

Depending on the purpose of the dose-finding trial – showing dose-response, identifying clinical relevance, selecting a target dose or estimating the dose response curve – the required number of patients varies greatly. We investigate how many patients are needed under the different scenarios, which turns out to be thousands of patients for anything beyond showing simple dose-response. This is generally not feasible in phase 2 trials, but a characterization of the dose-response is required from regulatory authorities.

Novel thinking is needed to speed up, and reduce the cost of, clinical development in phase 2. We outline routes to decrease the number of patients required by the use of biomarkers and novel surrogate endpoints. We show with extensive simulation that these approaches increase our chances of picking a suitable dose, and limit the risk of having to progress into phase 3 with multiple doses.

We exemplify these ongoing efforts with designs of AstraZeneca sponsored clinical trials of novel drugs aimed at reducing the frequency of severe exacerbations in patients with COPD or asthma.

3. Ann-Kristin Petersen (Bayer): "Evaluation of Interethnic Differences of Drug Response in Whites and Asians using the Meta-Analytic Prediction Method"

The drug exposure of an investigational medicinal product might be affected by the ethnical background of patients. This might result in the need for dose adjustments to ensure treatment efficacy. Therefore, the assessment of interethnic differences in pharmacokinetics plays an important role in clinical development. Often, this is done by comparing data of pharmacokinetic parameters of a single clinical study in e.g. Asian subjects with aggregated data from historical clinical studies in a reference group, e.g. White subjects. A simple approach comparing studies with similar designs conducted in different ethnicities, has the risk of false detection of differences induced through between-trial variability. The meta-analytic prediction method (MAP), which currently gains popularity in Bayesian analyses for deriving prior information, appears to be promising for the purpose of aggregation of historical data. The information about the pharmacokinetic parameter of the reference group is described by the predicted posterior distribution of the MAP calculated by MCMC sampling and taking the between-trial variability into account. Our goal is to quantify the interethnic differences of drug exposure for Whites and Asians for one of our investigational medicinal products using MAP. For this, the predicted posterior distribution for pharmacokinetic parameters of White subjects derived through MAP is compared with the posterior distribution of Asian subjects derived from one clinical study using MC sampling. We compare the results with those of traditional methods.

Wednesday:

Creative Designs: More Time for Something a Little Bit Different!

We are always being asked to be more innovative in our jobs as statisticians. This session aims to share three more examples of people thinking differently with the problems of clinical trials and analyses.

Speakers:

1. Cyrus Mehta (Cytel): "Adaptive Multi-Arm Multi-Stage Group Sequential Clinical Trials"

The statistical methodology for two arm group sequential clinical trials has been available for at least 35 years. The generalization to adaptive two-arm group sequential designs became available only in the last decade thanks to seminal papers by Lehhmer and Wassmer (1999), Cui, Hung and Wang (1999) and Müller and Schäfer (2000). The next stage of development is the generalization of these methods to multi-arm multi-stage (MAMS) group sequential trials. The statistical methodology already developed for the two-arm case can, in principle, be extended to MAMS designs. In practice, however, the formidable computational problems that must be overcome have inhibited making these methods accessible for realistic designs. We will discuss our recent work on overcoming these computational hurdles and will demonstrate the use of these methods for adaptive clinical trials that include early stopping, dropping of losers, and sample size re-estimation within the setting of adaptive dose selection and adaptive population enrichment. Comparisons will be made to alternative approaches that involve p-value combination and closed testing.

2. Pollyanna Hardy (National Perinatal Epidemiology Unit Clinical Trials Unit, University of Oxford): "The Coronis Trial: A Non-Regular Fractional Factorial Design. Why and How?"

Introduction: Caesarean section is one of the most commonly performed operations worldwide. A variety of surgical techniques for all elements of the operation are used. Many of them have not been rigorously evaluated in randomised controlled trials. The CORONIS Trial set out to simultaneously examine five elements of the caesarean section operation in seven low- to middle-income countries, using a novel adaptation of the factorial design.

Design: During the planning stages of CORONIS, five pairs of caesarean section techniques were agreed upon using a consensus process, for evaluation in a factorial design. It subsequently became apparent that, for pragmatic reasons, not all five pairs could be randomly allocated in all participating countries. Therefore, each participating site was assigned 3 of the 5 intervention pairs. In addition it was agreed that examining interactions was not important. This led to the adoption of a non-regular fractional factorial design, the first known use of its kind in a clinical trial setting. This presentation will explain how pragmatic considerations influenced the design of the CORONIS Trial and their implications on the sample size, randomisation, central monitoring, conduct and analysis.

Conclusions and recommendations: Our experience of using a complex trial design to deliver robust and reliable results was successful, but required a team of innovative researchers and clinicians. We would recommend undertaking such a design only with extreme caution.

References: 1. CORONIS Trial Collaborative Group. The CORONIS Trial. BMC Pregnancy and Childbirth 2007;7:24. 2. Mee RW. A Comprehensive Guide to Factorial Two-Level Experimentation. Springer: 2009.

3. Chris Harbron (Roche): "PIPE : A flexible, non-parametric model for dual agent dose escalation studies with visualisations and an interactive interface to facilitate adoption."

Background: With the increasing development of targeted drug combinations particularly of immunotherapies, dual agent trials are

becoming increasingly common. The superiority of model based designs over a 3+3 rule based design has been well established for testing single agents. With multiple agents, designs exploring the two-dimensional dosing space are required to identify the optimal benefit-risk dosing, as well as being understandable to clinicians with graphical visualisations and interactive interfaces to facilitate their adoption.

Method: PIPE is a novel curve-free (nonparametric) design for a dual-agent dose escalation trial assuming monotonicity. Estimation of the maximum tolerated dose contour is used to define the dose-escalation decision process by applying conjugate Bayesian inference, allowing rapid updating of the model and clinicians' prior beliefs or historical data to be easily incorporated. This is supplemented by a range of graphical outputs facilitating communication to clinicians and informed decision making throughout the study.

Results: The performance of the PIPE design is similar to that of other proposed models for dual agent dose escalation with gains when the observed toxicities do not fit the assumed models. We discuss improvements in the PIPE design to provide more flexibility and avoid the issues of rigidity that had been observed in some scenarios and describe graphical presentations and interactive software permitting clinicians to interact with the model, explore "what-if" scenarios and prepare "playbooks" of next steps from a range of potential outcomes in advance of the final data from a cohort, minimising delays for analysis and decision making within studies.

Dose Exposure Modelling

This session aims to show three diverse examples of Dose -> Exposure, Dose -> Exposure -> Response and Dose -> Response modelling. We hope the variety of this session and the engaging speakers will promote interesting ideas you can take back to your own projects and the opportunity for discussion. Please see the individual abstracts and we look forward to seeing you on the day.

1. Hanna Silber Baumann (Roche): "When target expression drives drug exposure – an example with the fusion protein CEA-IL2v"

Authors: Hanna E Silber Baumann1, Benjamin Ribba1, Hans-Peter Grimm2, Zhi-Xin R Xu3, Christophe Boetsch1, Valerie Cosson1

Affiliation: Roche Pharma Research and Early Development 1Clinical Pharmacology, Roche Innovation Center Basel, 2Pharmaceutical Sciences, Roche Innovation Center Basel, 3Clinical Pharmacology, Roche Innovation Center New York

For most drugs the pharmacokinetic (PK) properties are independent of the pharmacodynamic (PD) effects. However, for certain drugs, the PD effects elicited by the drug affects the PK properties. One example commonly seen with therapeutic monoclonal antibodies (mAb) is target mediated drug disposition (TMDD), which is a phenomenon in which a drug binds with high affinity to its pharmacological target (such as a receptor). When the target is not fully saturated this manifests as a non-linear disposition of drug with respect to concentration. For the example presented here, expansion of the target adds an additional layer of non-linearity which further impacts drug disposition over time. Characterization of TMDD behavior is of importance for designing dosage regimens and determining correlations between drug concentration and effect.

CEA-IL2v is a tumor targeted antibody-cytokine fusion protein with an IL2 variant which has been designed to overcome limitations of wild type IL2 (Proleukin). CEA-IL2v preferentially activates CD8+ and NK immune effector cells and is targeted to tumor types that express carcinoembryonic antigen (CEA). Non-linear elimination was observed with respect to concentration and time.

The model for CEA-IL2v included a TMDD component with the target as a latent (unobserved) variable. The model provided a quantitative semi-mechanistic framework to understand the interplay between PK and target expression over time. It has been used both in pre-clinic and in clinic to successfully explain the non-linear pharmacokinetic properties of CEA-IL2v elimination. The model captured the concentration-time profiles of CEA-IL2v well over multiple dose levels and treatment cycles. The knowledge gained through this analysis has supported development decisions on dose and regimen.

2. Tobias Mielke (ICON): "Two-Stage Dose Finding Designs using MCPMod - When to look and how to adapt"

Uncertainties in the dose-response relationship and in the magnitude of drug related effects pose challenges in designing dose-finding studies. The number of required patients and their optimal allocation depends on unknown parameters of the dose-response model. Misspecification of these parameters may lead to inefficient trial designs. Two-stage designs may increase the efficiency, utilizing interim information for an improved allocation of patients to the most informative doses. The interim timing and the considered method for the adaptation have a high impact on the operating characteristics of adaptive trial designs. Different adaptation rules in frequentistic adaptive two-stage designs will be introduced and their potential benefits and limitations will be discussed in a simulation study.

3. Jixan (Jason) Wang (Celgene): "Dose-exposure-response modeling and dose-escalation/adjustment"

Dose escalation and adjustment are common in clinical trials, including phase I dose-escalation trials and phase II-III trials with therapeutic dose monitoring (TDM) or dose reductions due to, but not limited to adverse events. Dose-exposure-response (DER) modeling can provide important information about the relationship between dose, exposure (PK), and response, and can also guide better dose escalation and adjustment. This presentation will consider two relevant topics. The first deals with the impact of exposure or response dependent dose escalation/adjustment on DER modeling. I will illustrate some common issues in routine assessments for DER relationships, such as dose-proportionality based on these trials, and recommend methods for DER modeling and model interpretation.

The second topic considers the use of DER modeling for better dose-escalation/adjustments. For phase I dose-escalation trials, I will show the advantage of using DER modeling to aid dose escalation over using empirical or dose-response model based algorithms, specifically for drug combinations and for drugs with prior PK knowledge. The use of joint modeling and sequential modeling for PK and response approaches will be compared, and the possibility of using patient level data will also be explored. I will also illustrate some caveats and potential issues when applying DER modeling based approaches for dose escalation. An outlook on finding personalized dosing strategies with DER modeling will also be presented. For illustration, this presentation will use numerical examples as well as a phase I trial design with hypothetical parameters reflecting real scenarios.

Risk Based Monitoring

Risk based monitoring (RBM) or statistical monitoring is a growth area in clinical trials. During the trial conduct, statistics is used to monitor risk factors, including data quality, compliance and patient safety. Today's presentations will focus on practical examples of RBM from statisticians with a variety of backgrounds and perspectives. Join us or risk missing out.

1. Marc Buyse (Chief Scientific Officer, IDDI & CluePoints): "Central statistical monitoring of clinical trials for better quality at lower cost"

Regulatory agencies encourage risk-based monitoring as a way to both improve data quality and streamline clinical trial costs. Risk-based

monitoring should be driven by indicators of the quality and performance of investigational sites. A common way to evaluate site performance is to monitor predefined metrics, often called "Key Risk Indicators" (KRIs). These typically include variables known to be relevant indicators of quality (e.g. accrual rate, frequency of adverse and serious adverse events, frequency of data queries and time taken to resolve them). A very informative complementary approach is Central Statistical Monitoring (CSM), which uses statistical tests to compare sites to each other and detect unusual data patterns that may be indicative of errors, misunderstandings, sloppiness, data fabrication or fraud. This talk will focus on the statistical methods used to implement a CSM software that performs as many statistical tests as possible on all available data collected in a clinical trial.

The software generates a high-dimensional matrix of P-values which are summarized in an overall score for each center. The score is analogous to an average P-value, with low scores indicating centers that are statistically most different from all others. A risk-based monitoring strategy can use this information, along with other performance indicators, to prioritize on-site monitoring for sites identified at higher risk. In multinational clinical trials using regional monitoring teams, the approach can also unveil data patterns that reflect unanticipated differences in patient management or other important aspects of trial conduct. Finally, when a trial is completed, the same approach can be used to check that the data are statistically consistent across all sites, regions or countries.

2. Alun Bedding (AstraZeneca): "Approaches to Detect and Preventing Fraud and Misconduct in Clinical Trials: A TransCelerate Review and Recommendations"

Fraud and misconduct at investigational sites in the pharmaceutical industry negatively affects the integrity of the clinical data and the validity of the trial results. Extreme cases may jeopardize the acceptance of study data. Despite this, sponsors have not consistently established advanced processes and systems to proactively detect and mitigate this type of risk.

A data set was created with simulated fraudulent data points. An independent statistical group was tasked with running a series of statistical tests to see if it could identify the fraudulent data. TransCelerate reviewed the results and will offer recommendations on how to effectively identify and manage certain risks to data integrity.

3. Amy Kirkwood (Cancer Research UK & UCL Cancer Trials Centre): "Central statistical monitoring in an academic clinical trials unit"

Within our academic clinical trial unit we follow a risk based approach to monitoring, avoiding, when possible, on-site source data verification which is an expensive activity with little evidence that it is worthwhile. Central statistical monitoring (CSM) has been suggested as a cheaper alternative where checks are performed centrally without the need to visit all sites.

We developed a suite of R-programs which could perform data checks at either a subject or site level using previously described methods or ones we developed. These aimed to find possible data errors such as outliers, incorrect dates, or anomalous data patterns; digit preference, values too close or too far from the means, unusual correlation structures, extreme variances which may indicate fraud or procedural errors and under-reporting of adverse events. The aim was to produce programs which would be quick and easy to apply and which would produce simple tables or easy-to-read figures. We will summarise the methods and, using examples from trials within our unit, show how they are implemented and that they can be easy to interpret.

We found CSM to be a worthwhile alternative to on-site data checking and may be used to limit the number of site visits by targeting only sites which are picked up by the programs. The methods can identify incorrect or unusual data for a trial subject, or centres where the data considered together are too different to other centres and therefore should be reviewed, possibly through an on-site monitoring visit.

Challenges in Pediatric Development: Kindermedizin not Child's Play

For nearly ten years now all new drugs have been required to have a paediatric investigation plan (PIP) in Europe and since even longer in the U.S. However in practice paediatric development is harder than it seems. During this session three statisticians will speak of their experiences with paediatric studies from a variety of perspectives. Please join us.

Speakers:

1. Lieven Kennes (University of Applied Sciences of Stralsund): "Challenges of paediatric clinical trials – practical experiences from a real case study"

In the last decade, paediatric clinical trials in drug development gained increasing attention, in particular due to specific requirements of the EU PDCO and US FDA. The planning, conducting and reporting of paediatric clinical trials differentiate considerably from adult clinical trials and yield a variety of ethical and operational challenges.

Since the beginning of 2015, Grünenthal GmbH is conducting a phase III, randomized, multi-site, double-blind, placebo-controlled, parallel group, multiple oral dose trial in children aged from birth to less than 18 years old. The study is designed to simultaneously fulfil different requirements of two regulatory agencies (EU PDCO and US FDA). The study design, specifically adapted to the paediatric population is presented. Statistical analyses of this trial, including a Bayesian analysis, are addressed. From this real study example, ethical, regulatory and operational challenges are shared together with their solutions.

2. Thomas Dumortier (Novartis): "Supporting a pediatric investigational plan using model-based extrapolation from adults"

Objective: To review a model-based extrapolation approach used for supporting a pediatric investigational plan (PIP) in solid organ transplantation (Tx).

Method: An extrapolation methodology was developed to bridge a recruitment gap in the planned sample size of the pediatric liver Tx study.

Results: After liver transplantation, the effect of time-varying drug concentration on the hazard of efficacy events (biopsy proven acute rejection, graft loss, or death) was estimated from the data of a Phase III study in adult patients by means of a time-to-event model; the drug concentration was predicted using a non-linear mixed effect model. The extrapolation concept was that the concentration-event relationship also holds in children. Under this assumption, the efficacy of patients treated similarly to those in the paediatric study could be predicted from the model. The extrapolation concept was then validated by checking if the predicted pediatric efficacy was in the range of the efficacy actually observed in the pediatric study. The concept was also externally validated using published pediatric information.

Conclusion: The extrapolation methodology supported evidence from the pediatric studies that efficacy was the same or better in children than adults, and a meta-analysis of key safety events identified

substantial differences between adults and children which corroborated the expected clinical differences between the populations. PK/PD modeling and meta-analysis provide tools for synthesizing evidence in the pediatric population and in comparison to the adult population, in line with the recent EMA concept paper on extrapolation (EMA 2013).

References: European Medicines Agency. Concept paper on extrapolation of efficacy and safety in medicine development. EMA/129698/2012 (2013)

3. Franz König (Medical University of Vienna): "An extrapolation framework to specify requirements for drug development in children"

A fully independent drug development programme to demonstrate efficacy in small populations such as children may not be feasible and/or ethical, especially if information on the efficacy of a drug is available from other sources. In a Bayesian framework and under the assumption of a successful drug development in adults, we determine the amount of additional evidence needed in children to achieve the same confidence for efficacy as in the adult population. To this end, we determine when the significance level for the test of efficacy in confirmatory trials in the target populations can be relaxed (and thus the sample size reduced) while maintaining the posterior confidence in effectiveness. An important parameter in this extrapolation framework is the so called scepticism factor that represents the Bayesian probability that a finding of efficacy in adults cannot be extrapolated to children. The framework is illustrated with an example.

Reference: Hlavin, G., Koenig, F., Male, C., Posch, M., & Bauer, P. (2016). Evidence, eminence and extrapolation. Early View. Statistics in Medicine. Free download from http://onlinelibrary.wiley.com/doi/10.1002/sim.6865/full.

Translating Signals to Outcomes: Challenges in Early Phase Alzheimer Development

1. Angelika Caputo (Novartis): "Multiplicity adjustment in a complex clinical trial in preclinical Alzheimer's disease"

By Angelika Caputo, Novartis Pharma AG, and Ron Thomas, University of California, San Diego

The Generation study (NCT02565511) aims to investigate the efficacy of two investigational compounds in comparison to respective placebo in participants at risk for the onset of clinical symptoms of Alzheimer's disease. The purpose of this study is to determine the effects of each of the two therapies given separately targeting amyloid on cognition, global clinical status, and underlying pathology. Cognitively unimpaired individuals with APOE4 homozygote genotype and age 60 to 75 years are selected as they represent a population at particularly high risk of progression to MCI due to AD and/or dementia due to AD. The trial will investigate hypotheses on two primary endpoints where success of the trial will be declared if at least one of the two endpoints reaches statistical significance. The strategy to adjust the type-one error rate is following a gatekeeping procedure based on the graphical approach to sequential rejective multiple test procedures (Bretz et al., 2009; Maurer et al., 2011).

Main design features of the trial will be presented with special focus on the approach to adjust for multiple testing, the strategy for the primary analysis, and on the role of the pooled placebo arm as a control group.

2. Nicola Voyle (King's College London): "Using Bayesian logistic regression analysis in the search for a multi-modal blood biomarker of Alzheimer's Disease"

It is thought that Alzheimer's Disease (AD) is caused by the aggregation of amyloid and tau proteins in the brain. In this study, we are searching for a blood biomarker of amyloid with the aim to enrich clinical trial populations for subjects with a high risk of developing AD or for amyloid positivity in a symptomatic population.

Single modality biomarkers (for example proteins, gene expression or metabolites) have shown some associations with amyloid. However, in general, replication has been poor. We hypothesise that by combing independent signal from multiple modalities we can create a better predictor of amyloid burden. Furthermore, we aim to increase the reliability of estimates for well-known risk factors for AD such as age and the APOE gene. There is an abundance of information available on the association of such risk factors with amyloid enabling us to incorporate this historical information in a Bayesian framework [1, 2]. We use Bayesian logistic regression analysis to incorporate informative priors for age and APOE based on large meta-analyses. Less is known about the associations between blood biomarkers and amyloid and so non-informative prior distributions are used for modelling. We use the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort alongside subjects from the EDAR [3] and DESCRIPA [4] studies to build and test models of dichotomised amyloid using age, APOE, polygenic risk scores, gene expression risk scores, proteins and metabolites.

This presentation will summarise the methods used and present some preliminary results.

References: [1] Jansen, W.J. et al. (2015) Prevalence of Cerebral Amyloid Pathology in Persons Without Dementia. Journal of the American Medical Association, 313 (19), 1924 – 1938.

[2] Ossenkoppele, R. et al. (2015) Prevalence of Amyloid PET Positivity in Dementia Syndromes. Journal of the American Medical Association, 313 (19), 1939 – 1949.

[3] Barnett, J. et al. (2010) Cognitive function and cognitive change in dementia, mild cognitive impairment, and healthy aging: The EDAR study. Alzheimer's and Dementia, 6 (4), S127.

[4] Visser, P.J. (2008) Development of Screening Guidelines and Clinical Criteria for Predementia Alzheimer's Disease. Neuroepidemiology, 30, 254 – 265

3. Trevor Smart (Eli Lilly): "Alzheimer's Disease should we jump, sink or swim through phase 2? How do different early phase designs address Alzheimer's issues?"

Alzheimer's disease is a key unmet medical need and the search is still on for disease modifying compounds. There are many issues unique to AD that need addressing in early phase development (up to Phase 3):

- Cognitive decline takes a long time, so for a proof of concept study, it needs to be a long study to be able to identify a reduction in the decline or the treatment difference will be small requiring large sample sizes.
- What population should we be treating, those later in progressing have a more rapid decline, but we may be too late to treat these patients, or do we choose patients with a much earlier diagnosis, but the progression is slower and hence a longer study is required.
- Can we use other endpoints, such as biomarkers instead of cognition for study endpoints or for identifying potential patients?

How with small studies can we gain enough confidence that the compound works, choose a population, dose and frequency of dose before going to phase 3? These and other issues will be discussed and alternative early phase options presented with simulations showing their advantages and disadvantages.

Breakout Session: Estimands in Practice

Defining the primary objective of a clinical trial in the presence of non-compliance or non-adherence to the assigned treatment is crucial for the choice of design, the statistical analysis and the interpretation of the results. This raises the need to clearly specify the primary estimand (i.e. “what is to be estimated”). In this break-out session, we will consider the choice of estimands and impact on design, trial conduct and statistical analyses for different clinical trial settings. A great chance to discuss this topic with your colleagues and pose any questions to the experts.

Regulatory Hot Topics Panel Session (Biosimilars, Extrapolation and Multiplicity)

This session will focus on topics which are currently of interest in the European regulatory setting. Starting with the statistical issues encountered in biosimilar applications such as whether ratios or differences should be used for the primary endpoint and whether the confidence intervals should be 95% or 90%. The next talk will move onto extrapolation and the EMA’s “draft reflection paper on extrapolation of efficacy and safety in paediatric medicine development”, with a special focus on the statistical aspects. The session will finish with a discussion of the new multiplicity guideline, the motivation for developing it and any changes that have been introduced from the original CHMP EMA Points to Consider document currently in use.

Speakers:

1. David Brown (MHRA): Abstract not provided

2. Norbert Benda (BfArM): “Regulatory Hot Topics: Multiplicity”

In 2002, the CHMP EMA Points to Consider (PtC) on Multiplicity Issues in Clinical Trials was adopted. Although the PtC has been proven to be useful for both, industry and regulators when planning and assessing confirmatory clinical trials, methodological advances have been made in more complex multiplicity settings relating to multiple sources of multiplicity, as different dose groups or treatment regimens, interim analyses, multiple endpoints, and different subgroups. In addition, regulatory requirements have been refined referring to the hypothesis framework in confirmatory clinical trials. Following the CHMP Concept Paper on the Need for a Guideline on Multiplicity Issues in Clinical Trials published in 2012, a new guideline on multiplicity has been drafted and is expected to be published soon. The presentation briefly outlines the regulatory principles related to multiplicity issues in drug approval and discusses changes in the new guideline document, the role of secondary endpoints and subgroup analyses, potential consistency and interpretational problems, multiplicity issues in estimation and how multiplicity procedures could be used to optimally support the risk benefit assessment.

3. Andrew Thomson (EMA): “Regulatory Hot Topics: Extrapolation”

In April 2016 the European Medicines Agency published the ***draft reflection paper on extrapolation of efficacy and safety in paediatric medicine development.*** The reflection paper was developed to communicate the current status of discussions on extrapolation of data and provides a framework for discussion or clarification particularly in areas where scientific knowledge is fast evolving or experience is limited. This talk will cover the key points of the reflection paper, and explain the drivers behind some of the concepts outlined. There will also be a focus on the statistical aspects covered in the paper, and on how statisticians can get involved in the process.

PSI Conference 2016: Collated Poster Abstracts

1. Sail on Silver Girl (Bridge over Troubled Waters)

Andy Kenwright (Roche)

Extrapolation to children is straight-forward, right? Get some adult efficacy and safety data, then some data in teenagers and work some PK/PD magic to “build a bridge”. Get a few younger patients’ safety data and you’re sorted. Rarely! When looking at sub-groups of patients within your target population e.g. Immune-compromised (IC) children in Virology, ensuring you have the right dose is a double edged sword. There is no evidence that neither a virus nor the investigational compound should act differently in IC patients, is there? No evidence that otherwise healthy (OH) children should behave differently to IC kids, nor that their dosing regimen should be different. Is there? Absence of evidence is no evidence at all.

The onus is on the Marketing Authorisation Holder to achieve the best dose for all patients regardless of age, weight and health status. Do you extrapolate from IC adults to IC children, from otherwise healthy children to IC children....what about mechanistic models, updated PKPD models. Where does it all fit? Finding a clear path to the right dosing regimen in a hard to recruit population, via a label enabling route with PDCO/FDA (keeping both content), could be a minefield. We discuss some scenarios and collaborations.

2. Stratified Randomization in Comparative Clinical Trials in Small Populations

Lukas Aguirre Dávila, Armin Koch (Hanover Medical School)

Among the stratified allocation methods argued in the literature stratified permuted block randomization and Pocock & Simon’s marginal procedure are most commonly referred to and applied in the design of clinical trials. In the ongoing discussion about statistical properties of different allocation methods, few recommendations are given about the choice of stratification variables. These are defined by clinical reasons and may have considerable impact on the behaviour of the discussed allocation methods. We investigate the robustness and sensitivity to varying conditions: While it is evident that the number of stratification factors is limited by terms of balance and accidental bias, these limitations are also inherent in other aspects - especially in clinical trials with small sample sizes.

3. Making the Most of What You’ve Got – Bayesian Methods in a Sensitive Population with a Rare Disease, Failing Standard Therapy

Efthymia Iliopoulou (Veramed), Peter Colman (UCB)

The Bayesian framework allows for probabilistic statements which are better understood by the clinical teams. It can also reduce the sample size and the study duration by leveraging historical information. Hence, Bayesian designs, meta-analysis and simulations are becoming increasingly important in clinical development programs, especially when studying rare diseases and sensitive, treatment resistant populations.

We considered a number of designs for a study in children, suffering from a rare disease, who have not responded to standard therapy. This first in children study would optionally be extended, with the intention of supporting regulatory approval. We explored different scenarios, varying the study population, treatment regimen and the decision rules, in order to assess the operating characteristics of the study. We present the statistical approaches that we considered, ranging from a basic frequentist comparison to Bayesian mixed effects meta-analysis, and the tools that we used for the assessment of the operating characteristics.

4. Adaptive Designs

Paul Constantin, David Shaw (GSK)

In consumer health trials it is useful to consider an interim analysis in order to check the validity of the assumptions made in the original sizing of the study (i.e: treatment effect and variability). Such adaptive designs can speed up the process of development or save resources in the case of futile studies. Predictive and conditional powers are a means of estimating the probability of obtaining a statistically significant result at the end of the study on the basis of the results of an unblinded interim assessment and/or prior beliefs and can be calculated for different final sample sizes. The sample size can be adjusted to improve the chance of obtaining a positive result at the end of the trial. Decisions about stopping, continuing or extending the trial will then be based on these projected inferences.

Spiegelhalter has given formulae for both conditional and predictive power calculations that assume the residual variability is known, hence depending upon large sample theory. However, in reality this may not be the case. For large sample sizes this becomes less of an issue, however with small sample sizes caution on this assumption needs to be taken. The aim of this paper is to compare Spiegelhalter’s formulae, simulation methods and possible other methods (The bootstrap etc) of calculating post-interim powers at an interim look. We will also investigate the accumulation of the type 1 error in simple adaptive situations where the potential for early stopping is considered.

5. Bayesian Predictive Probability – Theory and Application to an Oncology Trial

Nigel Baker (AstraZeneca)

MEDIOLA is an early phase oncology trial investigating the tolerability and efficacy of a combination treatment in four different indications. We have utilised the Bayesian predictive probability design of Lee and Liu (2008) to allow a flexible regular review plan. In the poster we will describe some background to the method, how we implemented it in the MEDIOLA study and some lessons learned from doing this.

Ref: Lee JJ, Liu DD., Clin Trials. 2008;5(2):93-106. doi: 10.1177/1740774508089279. A predictive probability design for phase II cancer clinical trials.

6. Recent Experiences with Implementing Bayesian Designs and Interim Analyses in Early-phase Drug Development

Maria Costa (GSK)

In the early stages of drug development, where the focus is in learning rather than confirming, the Bayesian inference paradigm offers intuitive probabilistic statements as an alternative to traditional hypothesis testing. One of the key challenges when designing a clinical trial using Bayesian methodology is the communication with clinical teams who may not be familiar with the concept. This talk will cover some recent experiences with implementing Bayesian decision rules in early phase clinical trials. In particular, how clinical teams were introduced to the concept of interim analysis using Bayesian predictive probabilities as a risk mitigation strategy, and which graphical tools proved useful to support decision-making.

7. Design Investigations in High Order Exploratory Designs

Paul Constantin, David Shaw (GSK)

Sample size calculations can be a technically demanding task in complex study designs that do not permit the use of sample size/ power calculation software. In this presentation we use three examples of complex designs with success criteria as a combination of superiority and non-inferiority (NI) comparisons to perform the sample size/power calculation.

The three designs that we will be discussing are (i) k-period 6 treatment incomplete cross over design, (ii) 4 period 4 treatments cross over design and (iii) a 3 by 2 factorial design. The success criteria for all three designs remained similar (as possible), defined as a combination of superiority and NI comparisons. Differences between these designs (and what they can investigate) will be discussed together with the way we simulated data for each of the designs. Clear objectives and definition of success criteria will allow identification of the key factors in studies and focus the design on the important 'must haves' from the trial.

Data were simulated for 5000 studies using SAS IML. Mean treatment difference vector, and variance-covariance matrix, were based on previous study data. The NI margin and criteria for superiority were pre-set.

In design (i) we explore the benefits of varying k (number of periods), in (ii) the benefits of a more focused approach to investigating 'must haves' alone and in (iii) the simultaneous exploration of two factors and their interaction.

8. Designing Multi-Arm Multi-Stage Trials with the R Package 'MAMS'

Philip Pallmann, Thomas Jaki, Dominic Magirr (Lancaster University)

Multi-arm multi-stage designs make clinical trials supremely flexible and efficient, especially in early phases when there is still some uncertainty about the most promising of several active treatments e.g., multiple drugs, multiple doses of the same drug, or multiple combinations of drugs. We propose a multi-stage extension of the classical Dunnett test where several treatment means are compared to a common reference, with strong control of the familywise type I error rate. Futile arms are dropped at interim, and as soon as any treatment is shown to be better than control, the trial can be stopped early for efficacy.

We introduce the R package 'MAMS' that makes designing such multi-arm multi-stage trials with Gaussian endpoints easy. The package implements a single-step and a step-down variant of the procedure, as well as an option to incorporate unplanned design modifications using the conditional error principle. Another convenient feature is that effect sizes are parameterised in a way that does not require any knowledge about the variance in advance. We illustrate with examples how 'MAMS' can be used to calculate sample sizes and group-sequential stopping boundaries, evaluate the operating characteristics of a design, and adjust for unforeseen changes.

9. Evaluation of Complete-visit and Split-visit Clinical Trial Designs to Measure Time to Onset of a Minimum Clinically Important Different Effect

Wilmar Igl, Aaron Dane, Ian Hirsch, Fredrik Öhrn, Monika Huhn, Caffe Burman, Alun Bedding, David Ruau, for the Trial Design Modelling Centre (AstraZeneca)

Background: A study was designed to determine the time to onset of a minimum clinically important different (MCID) effect of an investigational product. Various visit schedules were assessed to optimize the trial design.

Methods: A sample of 200 patients (100 in each of the new treatment and control groups) was simulated over 12 weeks. The generated data was based on an exponential model, but also used alternatives (e.g. EMAX) to evaluate the robustness of methods. Time to MCID values were estimated using a descriptive and a parametric method. The parametric method was based on an exponential time-response model. Initially a default visit schedule was proposed. Alternative visit schedules which varied the timing at the initial visits using a complete-visit or a split-visit design were explored. In the split-visit design different cohorts of patients were allocated to different visit schedules in order to have assessments across all days during the first two weeks of the study.

Results: In general, descriptive estimates were preferred, if the visit schedule was correctly, but the analysis model wrongly specified. Parametric estimates were preferred, if the (exponential) model was correct, but the visit schedule wrong. However, descriptive time to MCID estimates from split-visit schedule designs were among the most robust across all simulations (including non-exponential simulations).

Conclusions: The descriptive method with a split-visit schedule was superior to the other study schedules. However, limitations regarding additional organisational complexity and loss of power for between-group comparisons at split visits have to be taken into account.

10. Model-based Dose Escalation Designs in R with crmPack

Daniel Sabanés Bové, Wai Yin Yeung, Giuseppe Palermo, Thomas Jaki (Roche)

Model-based dose escalation designs have gained increasing interest due to the need for more efficient and informative Phase I trials. The wide-spread implementation of such designs has been hindered by the need for either licensing specialized commercial software (e.g. FACTS) or programming the design and simulations from scratch for each project. Therefore we developed the R-package "crmPack", which is now publicly available on CRAN. By providing a simple and unified object-oriented structure for model-based dose escalation designs, crmPack enables the use of standard designs, full flexibility to adapt designs to the project, and easy extension to new designs. The framework for the designs comprises the data structure, regression model and prior distribution specification, maximum increment rules, derivation of the next best dose, stopping rules for adaptive sample size, adaptive cohort sizes, and starting dose and dose grid specification. In addition to the classic continual reassessment method (CRM) and escalation with overdose control (EWOC) designs with possibly advanced prior specifications with e.g. mixtures and minimal informative priors, crmPack currently features dual-endpoint (safety and biomarker) designs and two-part (SAD followed by MAD part) designs. As crmPack is actively being developed further, this list will be extended.

We introduce crmPack by outlining the design framework and show how easy it is to specify a design, run simulations, and report dose recommendations during the trial.

11. Trimentum™: Developing Simulations for SPCD Trial Design

Samuel Ruddell (PPD)

Sequential Parallel Comparison Design (SPCD/ Trimentum™) is a novel two-stage design which improves the detection of an accurate therapeutic signal in indications with a high placebo response. Trials

that utilize SPCD benefit from running complex simulations to determine optimal study design. The poster introduces the process and technical challenges of developing such a simulator. Key goals are shown and different computing environments are investigated. The challenges and limitations associated with each approach are discussed as are the solutions developed. The final method of utilizing cloud based computing and the processes set up surrounding this are described.

12. Adaptive Designs for Medical Device Development

Scott Mollan, Abdallah Ennaji (ICON)

The FDA has recently released a draft guidance document regarding the use of adaptive designs for medical device trials, including premarket approval, 510(k), de novo, humanitarian device exemption and investigational device exemption submissions. The adoption of adaptive design for the development of a medical device offers a number of valuable benefits. These include improved device development efficiency, increased odds of pivotal trial success, shorter time-to-market, and greater decision-making opportunities to optimise investment in a portfolio of products.

Our poster provides insight into valuable adaptive designs for medical device development, requirements for implementation, and case studies of device adaptive design trials conducted by ICON. We describe the types of adaptive designs and associated operational challenges. The two case studies provide implementation details including the choice of the adaptive design strategy, the operational measures to maintain trial validity and integrity as well as the chosen technology solution. We also provide a summary of the outcome for each of the case studies and estimated benefit of the adaptive design strategy.

13. Optimising Adaptive Enrichment Designs

Thomas Burnett (University of Bath)

Choosing the trial population can be challenging when there is uncertainty about the treatment effect in multiple sub-populations. Adaptive enrichment designs allow the recruitment of the trial to be changed based on observations at an interim analysis. With possible adaptation of the trial design there are multiple hypotheses and multiple stages of the trial which must accounted for to ensure control of the familywise error rate. The decision made at the interim analysis will dictate the possible outcomes of the trial and so optimising the decision rule is an important consideration.

We use a Bayesian framework optimise within a class of decision rules of a particular form, where the possible actions are designed to protect the familywise error rate. Furthermore we may use these possible actions to find the Bayes' optimal decision for a particular trial. We compare properties of these optimised adaptive enrichment designs with those of fixed sampling designs.

14. A Novel Cross Phase Design in Oncology

Elizabeth Pilling; Ed Casson; Barbara Collins (Early Clinical Development Biometrics, AstraZeneca IMED)

A Phase I/II design for an Oncology study to explore the safety and efficacy of combination therapy. The study is designed with considerations to achieving accelerated regulatory approval and with flexibility to include optional study parts that may be triggered to answer specific clinical questions dependent on emerging study data, judged against pre-specified decision criteria. The design includes the following study parts: (1) Dose escalation to identify tolerated doses and a recommended phase II dose of the combination; (2) Single-arm expansion to further explore the safety, pharmacokinetics and anti-tumour effects of the combination; (3) Further single-arm extension to provide additional efficacy and safety data to give the opportunity to start a separate Phase II/III study; (4) Randomised, blinded two-arm Phase II part (within the same study); (5) Separate single-arm dose expansion to explore potential drug-drug interactions.

Decision criteria will be employed at milestones during the study to determine which parts of the study should be initiated. The Phase II part includes an administrative interim analysis for internal decision making only to give the opportunity to initiate Phase III planning and start up a Phase III trial early. The design properties will be described (including the decision criteria and risks of making an incorrect decision at the interim), as well as the experience of submitting the Protocol to FDA and MHRA and other learnings so far.

15. Assessment of Various Continual Reassessment Method Models for Dose-escalation Phase 1 Oncology Clinical Trials: Clinical Trial Data and Simulation Studies

Gareth James (Phastar), Stefan Symeonides (Edinburgh Cancer Centre), Jayne Marshall (AstraZeneca), Julia Young, Glen Clack (AstraZeneca)

Background: The continual reassessment method (CRM) is considered more efficient and ethical than standard methods for dose-escalation trials in oncology, but requires an underlying estimate of the dose-toxicity relationship ("prior skeleton") and there is limited guidance of what this should be when little is known about this association. Aim: To compare the CRM with different prior skeleton approaches and the 3+3 method in their ability to determine the true maximum tolerated dose (MTD) of various "true" dose-toxicity relationships.

Methods: We considered eight true dose-toxicity relationships, four based on real clinical trial data (AZD3514, AZD1208, AZD1480, AZD4877) and four theoretical. For each dose-toxicity relationship we conducted 1000 simulations and used the 3+3 method and the CRM with six prior skeletons to estimate the MTD. This allowed us to understand the effect of each prior skeleton and assess performance through the proportion of simulations where the MTD is correct, underestimated or overestimated.

Results: For each scenario, the CRM correctly identified the true MTD more frequently than the 3 +3 method. However, the CRM more frequently overestimated, and less frequently underestimated the MTD compared to the 3+3 method. The ability of each CRM approach to estimate the MTD varied considerably between different scenarios, and starting with a prior skeleton which matched the true dose-toxicity curve did not guarantee the best result.

Conclusion: The CRM outperformed the 3+3 method. Further work is needed to determine the optimum combination of dose-toxicity model and skeleton.

16. Impact of Simulation Parameters on a Bayesian Analysis using the Markov Chain Monte Carlo Method

Mircea Jivulescu (Cmed Clinical Services)

Our intention is to compare results of Bayesian analyses using the Markov Chain Monte Carlo (MCMC) simulation method. A Bayesian binomial model using dummy data with similar characteristics as real clinical data was considered, comprising patients distributed across four active treatment groups and one placebo group. The analysis was based on response rates, calculated as the difference between active treatment arms and placebo. The prior distributions assigned were non-informative for the active treatment groups and informative for the placebo group. Different burn-in sizes, number of simulations and thinning rates were used, resulting in a large number of different scenarios and sets of results. The results were obtained using the PROC MCMC procedure in SAS with random-walk Metropolis sampling algorithm. The diagnostic and posterior information, the convergence of the Markov Chains resulting from running the analyses were assessed and the results (response rates, credible intervals and probabilities) of the analyses were compared.

17. A Bayesian Network Approach to Analyse 16S rRNA Bacteria Interactions in Infants Participated in a Nutritional Intervention Study. A Case-study on the Application of Bayesian Models for Integrative Data Research

Bianca Papi (OCS Consulting B.V.)

Background: The human gut microbiome plays an influential role in maintaining human health, and it is a potential target for the prevention and treatment of disease. Our gut is inhabited by different types of bacteria that naturally interact within each other. Understanding how bacteria are linked to each other and how different nutrition influences these links is important to preventing some diseases in the human body.

Aim: The primary objective of this study is to visualise probabilistic relationships among 16S rRNA genes that belong to different bacteria and their link to infants' crying. Microbial composition has been derived from stool samples of 111 infants randomly assigned to three tests products and one control group.

Methodology: A Bayesian Network has been applied to visualise probabilistic relationships among 16S rRNA genes that belong to different bacteria and their link to infants' crying. Four different algorithms have been implemented in order to learn the structure of the network and K-fold cross validation has been used to study the quality and robustness of the methodology.

Conclusions: The Bayesian Network, implemented through Hill-Climbing with BIC score, efficiently describes the interactions between bacteria in the analysed infants. It can be used, in integrative data research, to predict causal relationship between variables, especially when some biological knowledge is available a priori.

18. Haematology-Oncology Biomarker Story: An Example of Exploratory Analyses to Help Drug Development

Guiyuan Lei (Roche)

The talk will focus on the statistical challenges and key learnings/ recommendations of exploratory biomarker analyses, and how analysing historical biomarker data can help drug development. In this example, we retrospectively assessed the prevalence and prognostic value of BCL2 and other markers in patients from MAIN, a Phase III trial that evaluated bevacizumab plus R-CHOP in frontline, CD20-positive DLBCL. The main challenge for this exploratory biomarker analysis was the issue of multiplicity due to testing biomarker effects in different assays, cut-offs and subgroups. Statisticians played a key role in negotiating the primary set of biomarkers among many exploratory analyses by working closely with biomarker scientists. We will describe how hypotheses generated from those analyses are important for new drug development and to be validated in new clinical trials.

We will share following key learnings/recommendations from this biomarker story:

- Analyses of historical data (across projects) enable to evaluate hypothesis from research to gain more confidence
- It is critical to specify clear objectives/analyses even for retrospective exploratory analyses
- Multiplicity (false positive): to avoid extensive fishing of biomarkers, specify primary biomarker based on good scientific/biology knowledge, pre-clinical hypothesis or literature
- Capture as much as possible of the knowledge of the biomarker scientists to build into the analysis plan even when they didn't realize the importance of some of their existing knowledge

19. Integrative Modelling of Experimental Medicine Clinical Trial Data Shows That Engagement of an Investigational Monoclonal Antibody to its Target Predicts Clinically Relevant Biological Effects

Fabio Rigat, Stefano Zamuner, Kirsty Hicks (GSK)

Experimental medicine (EM) clinical trials are "skinny", in that the responses of relatively few individuals to therapeutic or prophylactic interventions are measured in depth using state-of-the-art bio-analytical techniques. One specific objective in EM is to establish whether the engagement of an investigational drug to its target (TE) predicts changes in the biological mediators of clinically relevant effects. Interpretation of these EM trials data requires integrative models of multiple bioassay readouts to robustly relate one independent variable to many dependent variables in presence of multiple sources of measurement error.

Here we show that discriminant analysis offers a Bayes-optimal tool for such robust integrative analysis of EM data when measurement error distributions are derived from bioassay validation experiments. Specifically, a dichotomised TE is shown to predict changes in leukocyte classes mediating epithelial inflammation after infusion of an investigational drug in healthy volunteers. Data uncertainty is propagated through simulation from the measurement error distributions and predictive accuracy is assessed using cross validation. Also, the Kullback-Leibler discrepancy (KL) between the distributions of all leukocytes across the two TE classes is shown to dominate the KLs measured for each of its margins, indicating that TE affects the whole population of leukocytes rather than one specific cell type. This result contributes to explaining the drug's immunological effects as well as posing the statistical bases for leukocyte data to be considered for defining surrogate endpoints.

20. A Bayesian Framework for Identifying Placebo Responders in Clinical Trials

Reagan Rose (Harvard University),
Rob Kessels (Emotional Brain/Utrecht University)

The placebo serves as an indispensable control in randomized clinical trials and allows us to draw causal conclusions about the effect of drugs. However, in trials with a subjective primary outcome where the expected placebo effect is large, the placebo can become more of a confounder than a control. To elucidate the true drug effect in this setting requires a better understanding of the placebo response and, subsequently, more careful statistical considerations of how to appropriately model and analyze such data.

We develop a flexible Bayesian framework for analysis of the placebo response using the Neyman-Rubin model of causal inference, which can be applied post-hoc to clinical trial data in order to obtain better estimates of the drug effect. In this framework, we propose that all units in a clinical trial can be classified as either placebo “responders” or “non-responders” and that the observed outcomes for each unit can be viewed as manifestations of an underlying distribution that is dependent on this missing “responder” status. Viewing the marginal distribution of outcomes as a finite mixture model, we then implement missing data methods via Markov Chain Monte Carlo (MCMC) in order to generate fully-Bayesian posterior probabilities that each unit is a placebo “responder” given their observed outcome. This general framework is extended to incorporate pre-treatment covariates, allowing for more flexible modelling. We illustrate this approach through an example with clinical trial data for treating female sexual dysfunction.

21. Bayesian Sensitivity Analysis for Drop-out in Parametric Time to Event Models Using Informative Missingness Hazard Ratio

Ákos Ferenc Pap (Bayer Pharma AG)

Introduction: Patients drop out from a clinical trial without having an event of interest and without information available afterwards. Sensitivity analysis should be performed to explore how an increased hazard of the event after drop out could influence the estimated cumulative risks by treatment group at the planned end of observation.

Method: Drop out subjects are assumed as having increased hazard function of the event after dropping out (h_dout[t]) as multiple of the hazard at the time point of drop out (h_ninf[t]). This factor can be named as informative missingness hazard ratio (IMHR; cf. 1, 2) and h_dout= h_ninf*IMHR. IMHR can be different by treatment group and by other factors. The expected cumulative hazard H_i is calculated for each patient i as the sum of cumulative hazard up to drop out and cumulative hazard after drop out (pattern mixture). Based on the means of cumulative hazards the mean cumulative risks by treatment group are estimated at the planned end of observation. The analyses are done with a Bayesian MCMC method using informative priors for the IMHRs.

Example: Event rates are in the experimental and control groups 5/100 and 15/100 respectively, the drop out rates are 10%. The risk ratio is 0.40 (95% CrI: 0.11 – 0.86) from a parametric Weibull model assuming non-informative censoring. The risk ratio applying larger IMHR for the experimental group (8 vs. 2) is 0.51 (95% CrI: 0.14 – 1.08).

References:

Higgins et al. Clin Trials. 2008; 5: 225–239.
Zhao et al. Biopharm Stat. 2014; 24:229-53.

22. A Simulation Study to Compare Re-current Event Methods with Time to First Event Analyses: Benefit vs. Complexity

Helen Millns, Tal Otiker, Yansong Cheng (GSK)

As part of the design of a parallel group, placebo controlled clinical outcomes trial, the methodology for analysing a primary endpoint which is a composite of several events is investigated via simulation. The conventional analysis is a time to first event analysis but this does not take into account the patients experience after the first event, when they may have further events of the same type or a different type. A range of recurrent event methodologies including the Prentice, Williams, Peterson (PWP), Wei, Lin, Weissfeld (WLW), Anderson-Gill, negative binomial and unmatched win ratio analyses were investigated and the pros and cons of each of these will be discussed.

Simulations were performed to determine the power of the different methodologies. Simulating recurrent event data from a composite endpoint requires a substantial number of complex assumptions e.g. the event rates and hazard ratios for each component and each recurrence of the event, and the correlation between different events. For this trial, which is predicted to have a high event rate, the simulations showed that all methods gave a small increase in power compared to a time to first event, except WLW which had a small decrease in power. The impact of changing the proportion of patients with each type of event, and the impact of delayed treatment response or early treatment withdrawal will also be discussed. The benefit of the recurrent event methods for this primary analysis was not considered worthwhile given the additional complications and limited increase in power.

23. Calculating and Applying Restricted Mean Survival Time (RMST) for Assessing Survival in the Absence of Proportional Hazards

Jonathan Wessen (AstraZeneca)

In randomised clinical trials with a right-censored time-to-event outcome the hazard ratio is often used to compare the efficacy of two or more treatments, however the hazard ratio is only valid when the proportional hazards (PH) assumption is satisfied. The absence of PH means that the hazard ratio between two treatments will vary depending on the time. One solution to this might be to consider an average hazard ratio; however its results can be misleading and therefore are potentially not a useful summary statistic.

The Restricted Mean Survival Time (RMST) has been proposed as a better summary statistic for when the PH assumption has been violated; Royston P et al. provide evidence in favour of its use as a primary endpoint in the absence of PH, and also suggest that it could be a useful secondary measure even when the PH assumption is satisfied. There are various different ways to calculate RMST which include calculating the area under the KM curve, Pseudovalues, and using a flexible parametric model; each method has its own strengths and limitations. Practical applications of these methods and their differences will be explored within this poster. Results will be illustrated graphically where possible and will be accompanied by code snippets.

24. The "Probability of Being in Response function" (PBRF) for Oncology Trials – Easier Said Than Programmed

Abram Graham (Quintiles)

In oncology clinical trials, the proportion of patients who respond to treatment, and the duration of their response before disease progression, are often of interest. The Probability of Being in Response function (PBRF) was first proposed by Temkin (1978) to simultaneously present the time to response and subsequent failure (i.e. progression) in a graphical way. This function, when plotted against time from the start of study medication, provides a visual method for comparing the frequency and duration of patient response between different treatment arms. A formal comparison between treatment arms based purely on duration of response, however, is likely to be biased (as groups are defined by the outcome of response, which is clearly likely to be treatment-related). Therefore, the PBRF was also adopted by Ellis et al (2008) to estimate the Expected Duration of Response (EDoR), in order to obtain a formal, unbiased comparison between treatment arms. However, despite the availability of literature on the subject, when it comes to practically programming the PBRF in statistical software such as SAS, available guidance is limited. We describe here our approach to developing the PBRF graph using SAS, and include the key SAS code you would need to apply to your own oncology trials. References:

Temkin NR. An analysis for transient states with application to tumour shrinkage. Biometrics 1978;34:571–80.

Ellis S, Carroll KJ, Pemberton K. Analysis of duration of response in oncology trials.

25. A Comparison of Agreement Between Endpoints with Respect to Decision Making for Overall Survival

Jayne Marshall (AstraZeneca), Lynsey Womersley (PHASTAR)

In Oncology clinical trials, one aim amongst many is to demonstrate improvement in life expectancy. Trials to evaluate overall survival can take a long time and it is in the interest of patients and advances in medicine to be able to make good decisions about the clinical development plan in light of study outcomes. Such decisions allow drug development to be continually working towards positive outcomes for patients.

Within clinical development, early decision making plays a critical part in contributing getting medicines to patients earlier. For outcomes that can take a long time to observe, such as survival time, it is of interest to understand whether there are other outcome measures that result in an early decision that is consistent with the decision based on the final overall survival analysis, using prespecified decision criteria. One such endpoint explored is the survival rate. Through simulations, this poster explores the question: Can a survival rate endpoint be used for early decision making when the final analysis endpoint is overall survival?

26. Assessing Minimal Residual Disease (MRD) as a Predictor of Progression-Free Survival (PFS) in Three Phase III Studies of Chronic Lymphocytic Leukaemia (CLL)

Natalie Dimier, Carol Ward, Paul Delmar (Roche)

The standard primary endpoint in clinical trials of chronic lymphocytic leukemia (CLL) is progression-free survival (PFS). With median PFS over 5 years, trials take many years to complete and are increasingly burdensome for the patients enrolled. Valid surrogate endpoints are urgently needed to reduce trial duration, thereby accelerating drug development. Patients who achieve minimal residual disease (MRD) of <1 clonal cell/10,000 leukocytes in peripheral blood at the end of treatment have been shown to experience significantly improved PFS. Together with the reduced observation time of MRD, this provides rationale for exploration of MRD response as a surrogate endpoint for PFS in this setting.

Based on three Phase 3 clinical trials (1203 patients), an evaluation of surrogacy was conducted. Prentice criteria were applied to each study, and a weighted linear regression model was used to assess whether the log relative-risk of MRD response could be used to predict the PFS hazard ratio. Results of the meta-regression show a significant association between treatment effects (for each unit increase in MRD log relative risk, the log of PFS HR decreases by -0.216; 95% CI, -0.358 to -0.074; p=0.005).

A surrogate endpoint should provide prognostic value for the specific clinical outcome, as well as evidence that treatment effect on the surrogate endpoint reliably predicts treatment effect on the clinical outcome. Our model suggests that treatment effect on PFS can be predicted by treatment effect on MRD response, and supports the use of MRD as a surrogate for PFS in patients with CLL.

27. A Flowgraph Model for Estimating Discrete Gap Times of Need-based Treatments

Lillian Yau (Boehringer Ingelheim GmBH), Ekkehard Glimm (Novartis AG)

Pro re nata, or PRN, is a need-based treatment regimen where patients receive medication only when symptoms occur. Examples are the use of sleeping aids or allergy relieves. Such a PRN regimen was implemented in a recently completed randomized phase III trial. Of interests are the number of consecutive treatments required, and the gap time between these treatment periods. We introduce a truncated geometric distribution for these discrete waiting times. To account for the possible dependency of the treatment and non-treatment cycles, we estimate the quantities of interest via a flowgraph model. Flowgraph models are a systematic procedure for inference and prediction of complex systems. The generality of the approach allows for great flexibility when modelling multi-state semi-Markov processes. In this presentation we illustrate how the method can seamlessly integrate with existing, familiar model building techniques.

28. Model Based Network Meta-Analysis: A Framework for Evidence Synthesis of Dose-response Models in Randomised Controlled Trials

D Mawdsley (University of Bristol), M Bennetts (Pfizer), S Dias (University of Bristol), M Boucher(Pfizer), NJ Welton (University of Bristol)

Model based meta-analysis (MBMA) has become an increasingly important tool in drug development since it was first proposed by Mandema et al. (2005). By including a dose and/or time response model in the meta-analysis it is possible to compare treatments at doses and/or times that have not been directly compared in head-to-head trials, potentially reducing the number of trials required, allowing the competitive landscape to be surveyed, and reducing the risk of late stage trial failure.

Network Meta-Analysis (NMA) is increasingly being used by reimbursement agencies in their policy decisions, where the focus is on making decisions about the relative efficacy and cost-effectiveness based on late stage state and post filing RCT evidence. However NMAs typically either consider different doses completely independently or lump them together, with few examples of models for dose. Methods of measuring model fit and the consistency of direct and indirect evidence within an NMA are, however, more developed than in MBMA.

We propose a Bayesian Model Based Network Meta-Analysis (MBNMA) framework, which combines both approaches. It combines evidence from randomised controlled trials (RCTs), comparing multiple treatments (agent and dose combinations), respects randomisation in the included RCTs, allows estimation and prediction of relative effects for multiple agents across a range of doses, uses plausible physiological dose-response models, and allows the assessment of model fit and evidence consistency.

29. Standardized Mean Differences in Network Meta-Analysis

Kelly Fleetwood, Daniel James, Ann Yellowlees (Quantics Consulting Ltd)

In meta-analysis (MA) and network meta-analysis (NMA), a single outcome may be measured on different scales. In some cases it may be possible to include all of the measures in a single MA or NMA by re-scaling the measurements, for example, weight measurements in stones could easily be converted to kilograms. However, re-scaling is not always possible; for example, there may be no validated method for conversion between different psychometric scales.In such cases, where re-scaling is not possible, standardized mean difference (SMD) approaches are often applied. The advantage of these methods is that a single analysis can incorporate all of the information about an outcome. However, these approaches require the assumption that any variability in the standard deviations reflects differences in the measurement scale rather than differences in the patient population. In practice, it is common for patient-to-patient variability to differ between trials and SMD approaches may not always be appropriate [1]. In this poster, we review the use of SMD approaches in the context of NMA. We illustrate our ideas with a case study based on a recently published NMA. For the case study, NMAs based on the raw data and the SMDs lead to very different conclusions and we investigate the underlying cause of the discrepancy. We conclude by providing some guidance on the use of SMD approaches in NMA.

[1] Ades, A.E., et al., Simultaneous synthesis of treatment effects and mapping to a common scale: an alternative to standardisation. Res Synth Methods, 2015. 6(1): p. 96-107.

30.Bayesian Extrapolation with the Help of Meta-analytic Methods

Kristina Weber, Armin Koch (Medical School Hannover)

In the planning of clinical trials, historical information based on other studies is always indirectly incorporated in the study design, especially when calculating the sample size of a new trial. Bayesian analyses offer an opportunity to use existing information not only in the design but also in the analysis, since the Bayesian philosophy allows evaluating the outcome of a new experiment “in the light of” prior knowledge. This methodology is often applied in situations with limited options to recruit patients into studies.

A special case is the use of Bayesian analyses in the context of extrapolating evidence from adult to paediatric trials. We introduce two motivational examples that represent cases with a different amount of already available adult information and newly observed paediatric data. These examples are then analysed with different meta-analysis based methods. In both the Bayesian meta-analyses and the Bayesian meta-analytic predictive approaches, it is necessary to specify prior distributions. This is either possible in a preferably non-informative way or by using the already existing adult information. We will in particular focus on the definition of heterogeneity prior distributions which are in case of only two studies never non-informative.

We see that the conclusion on the treatment effect is very much dependent on the analysis approach and prior distribution used. It is especially dependent on the assumptions of heterogeneity. A careful review of available analysis methods and their implications is necessary, especially in the context of Bayesian extrapolation.

31. An Application of the Estimands Approach for Assessing the Effect of a New Therapy in Idiopathic Pulmonary Fibrosis

Florence Le Maulf, Inga Tschoepe, Mannaig Girard (Boehringer-Ingelheim)

Background: The INPULSIS™ trials were two replicate, 52-week, randomized, double-blind, placebo-controlled Phase III trials that investigated the efficacy and safety of nintedanib in 1066 patients with idiopathic pulmonary fibrosis. The primary endpoint was the annual rate of decline in forced vital capacity (FVC). The estimand of main interest used all available FVC values from baseline to week 52 including FVC measurements obtained after premature treatment discontinuation. The statistical model was a random coefficient regression model with random slopes and intercepts. Missing data were not imputed for the primary analysis. The model allowed for missing data, assuming that they were missing at random.

Methods: Sensitivity analyses assessed varying assumptions for the handling of missing data at week 52, using different estimands. Multiple imputation methods were performed with different assumptions on the persistence of treatment effect in patients who withdrew from the trial prematurely.

Results: In both trials, all sensitivity analyses were consistent with the primary analysis of the annual rate of decline in FVC, confirming superiority of nintedanib versus placebo.

Conclusion: The use of a range of estimands to assess the robustness of the primary endpoint results was pre-specified following discussions with the FDA. This approach led to approval of the drug in major countries. No questions were raised by the regulatory agencies about the primary endpoint results and the missing data handling.

32. Bayesian Approach Based on Multiple Imputation Method as Compared to Classical Approaches to Handling Missing Data in Clinical Trials

Izabella Toth (Cmed Clinical Services)

Missing data is a frequent issue in data collected during clinical trials determined by different and various reasons. As ignoring missing data is not an acceptable option, some well-known classical approaches have been used though time to meet regulatory requirements. All classical approaches are based on single imputation methods, each having some specific advantages and disadvantages. A multiple imputation method based on the Bayesian approach seems be becoming a more popular method to handle missing data. The multiple imputation method is based on Markov Chain Monte Carlo and is implemented in SAS’s PROC MI and MIANALYZE. Multiple imputation methods generate multiple copies of the original data set

by replacing missing values using an appropriate stochastic model. The complete sets are then analysed and the different parameter estimates across the datasets are finally combined to produce a unique point estimate and standard error taking into account the uncertainty of the imputation process. This method provides valid statistical inferences that reflect the uncertainty due to missing data. We aim to investigate the impact of various methods used to handle missing data on inferences. Special focus will be considered for results obtained with the classical well-known Last Observation Carried Forward (LOCF) method as compared to the multiple imputation method based on Bayesian approach results. We will present also a set of summaries ignoring missing data, to assess the impact of missing data on the results. Different graphics will be presented to compare the obtained results with different methods to handle missing data.

33. A Tipping Point Analysis to Assess Sensitivity to Missing Data for a Time-to-event Endpoint

Abellan JJ, Russell C, Hammer A, Oldham M, Best N (GSK)

We present a tipping point sensitivity analysis to assess robustness to missing data of the results of a time-to-event analysis, using recently developed methods to handle informative censoring (Jackson et al 2014). This analysis was requested by the FDA in the context of a Phase III, randomized, double-blind, parallel-arm, placebo-controlled clinical trial with ~800 subjects. The original analysis was a Cox proportional hazards model (CPHm) to compare the efficacy of the experimental drug versus placebo. This analysis implicitly assumed censoring-at-random for the ~12% of subjects who discontinued. For the sensitivity analysis we considered censoring-not-at-random assumptions. Event rates for censored subjects were varied incrementally from an infinite event rate (worst case scenario where all dropouts experience the event) to a zero event rate (best case scenario where no dropout experiences the event). Event rates were varied independently for each arm. Using multiple imputation, time-to-event was imputed for censored subjects for each combination of assumed event rates in the two arms.

We fitted a CPHm to each of imputed dataset and combined the results using Rubin's rules (Rubin, 1987). Results were summarized in a tile plot (Liublinska and Rubin, 2014) showing how the estimated hazard ratio and associated statistical significance changed as the assumed event rates vary in each arm. By determining the change in the event rate in dropouts needed to tip the p-value above 0.05, and assessing how far the original results were from this tipping point, we concluded that the original results were robust to plausible missing data assumptions.

34. Identifying Unobserved Patient Subgroups and Estimating their Prognostic Relevance

Dan Lythgoe (PHASTAR)

In clinical trials we often estimate an 'average' treatment effect and then use interactions to test whether a treatment effect varies across known subgroups. However, it's plausible that our study population is composed of unobserved subgroups of patients for whom the treatment effect may vary. Larsen (2004) introduced a joint latent class and survival model (JLCSM) for identifying latent subgroups in binary covariate data and simultaneously estimating their prognostic relevance. We expand the JLCSM to include mixed data types and use simulated data to evaluate the usefulness of these models to identify prognostically relevant, unobserved patient subgroups. Finally we apply the JLCSM to a well-known prostate cancer data set.

References:

Larsen, K. (2004). Joint Analysis of Time-to-Event and Multiple Binary Indicators of Latent Classes. Biometrics (60) 85-92.

35. Subgroup Identification in Personalized Treatment of Alcohol Dependence

Hou, J., Seneviratne, C., Su, X., Taylor, J., Johnson, B., Wang, X.-Q., Zhang, H., Kranzler, H. R., Kang, J. and Liu, L. (Northwestern University, Chicago)

Identification of patient subgroups to enhance treatment effects is an important topic in personalized (or tailored) alcohol treatment. Recently, several recursive partitioning methods have been proposed to identify subgroups benefiting from treatment. These novel data mining methods help to address the limitations of traditional regression-based methods that focus on interactions. We propose an exploratory approach, using recursive partitioning methods, for example, interaction trees (IT) and virtual twins (VT), to flexibly identify subgroups in which the treatment effect is likely to be large. We apply these tree-based methods to a pharmacogenetic trial of ondansetron. Our methods identified several subgroups based on patients' genetic and other prognostic covariates. Overall, the VT subgroup achieved a good balance between the treatment effect and the group size. In conclusions, our data mining approach is a valid exploratory method to identify a sufficiently large subgroup of subjects that is likely to receive benefit from treatment in an alcohol dependence pharmacotherapy trial. Our results provide new insights into the heterogeneous nature of alcohol dependence and could help clinicians to tailor treatment to the biological profile of individual patients, thereby achieving better treatment outcomes.

36. A la CART or Table d'hote? Variable Selection in Rheumatoid Arthritis

Paul Mahoney (Roche Products Ltd), Steven Julious (University of Sheffield), Mike Campbell (Roche Products Ltd)

Background: During the clinical development of a new compound for the treatment of Rheumatoid Arthritis we had the opportunity to explore the data accumulated from 4 previous clinical programs covering more than 10 years of RA clinical development. We were interested to understand how selection of patients might influence patient outcome, specifically which patient characteristics measurable at screening and baseline would be important either in a univariate or multivariate fashion. With scores of biomarkers of which only a few may be predictors, this poster describes simulations to compare the performance or Classification and Regression Trees (CART) and Logistic Regression in variable selection.

Methods: Following development of a simulation protocol, simulations were conducted in SAS and R. Scenarios were generated on multiple response rates (resembling ACR20, ACR50, ACR70), correlations (from 0 to 0.90 correlated with response), sample sizes (n=100, 1,000, 10,000) and missing value assumptions for 144 separate scenarios. Each scenario was repeated 100 times creating 14,400 datasets with almost 4 million simulated patients.

Results: CART outperforms logistic regression in most scenarios. In an ideal situation where there are no missing values and all explanatory variables are continuous, throughout 144 variations of this scenario CART clearly outperforms logistic regression, particularly in the case of high subject numbers (n=10,000). Improvement of the performance of logistic regression can be improved by selecting significance levels for entry and staying set at low values.

Conclusions: Within the scope of the simulations, CART outperforms logistic regression in most scenarios.

37. Assessing Equivalence in Rheumatoid Arthritis

Niccolo' Bassani (Quanticate)

In our rheumatoid arthritis (RA) case study, interest lies in setting equivalence criteria for the efficacy of a bio-similar relative to the standard treatment.

The response to RA treatment is commonly assessed using the American College of Rheumatology Arthritis composite responder score for a 20% improvement (ACR20). The most common approach focuses on comparing proportion of responders at the last time-point using either an approximate or an exact test. Reeve et al. (Ther Innov Regul Sci 2013) fit an exponential time-response model to the ACR20 profile of proportion of responders of each treatment across time, and starting from this model Choe et al. (Ann Rheum Dis 2015) calculate a summary statistic "2-norm" for the difference between the fitted treatment curves (f and g) based on the square root of the integral of (f-g)2.

We compare existing approaches and explore potential alternatives in a simulation framework under various scenarios, in order to make recommendations for a suitable difference measure and equivalence criterion.

38. Estimating Marginal Distributions at Endpoint when only Baseline and Relative Improvement is Available: A Two-dimensional Probability Model for the Psoriasis Area and Severity Index (PASI) to Transform Relative PASI Improvements of ≥75%, ≥90% and ≥100% (PASI75, PASI90 and PASI100) into Proportions of Patients Reaching Absolute PASI ≤1, 3 or 5

Helmut Petto, Alexander Schacht (Eli Lilly)

The PASI (range: 0–72) is a standard measure used in studies with psoriatic patients. Studies usually include only moderate to severely ill patients (PASI≥12 at baseline). Publications contain only baseline mean (standard deviation [SD]), and proportions of patients achieving PASI75, PASI90 or PASI100 at endpoint. Recently, the focus has changed to also report proportions of patients reaching absolute PASI values of □1, 3 or 5 at endpoint. As it is very difficult to compare these absolute PASI changes with existing literature, we developed a tool to estimate absolute PASI levels from aggregated results. The tool is based on a latent two-dimensional normal distribution for baseline and endpoint PASI with a truncation to allow for the baseline inclusion criterion.

Baseline and follow-up values are logit-transformed, giving a truncated version of a two-dimensional logistic normal distribution, with values of the logit-normal marginals restricted between 0 and 72. A given PASI100 can be added to the follow-up marginal, so the PASI 1, 3, 5 probabilities can be derived from the resulting mixed distribution. Expected baseline mean (SD) and PASI75, PASI90 and PASI100 values can be derived from this model, so that by minimizing the distance to given aggregated values, model parameters and then PASI 1, 3, 5 proportions can be estimated. We fitted a least-squares model to aggregated results from 3 phase III studies, with known PASI ≤1, 3, 5 proportions. The predictions represented the real results very well. However one limitation is that only effective treatments can be analyzed.

39. Principal Component Analysis and its Merits

Ingrid Franklin (Veramed), Emma Jones (Veramed), Margaret Jones (UCB)

Results from small exploratory studies are often interpreted using only descriptive statistics because sample sizes are too low to perform any meaningful formal statistical testing. It can be difficult to extract true patterns for the overall population from the available data, especially in cases where there is a larger number of exploratory variables than there are subjects. Principle components analysis (PCA) is a form of multivariate modelling that can help to simplify interpretation of such results by reducing the dimensionality of the dataset. Using PCA, it may be possible to 'tease' out underlying and more intricate patterns that could be missed by simply observing summary statistics or plots, and when the two methods are used in tandem a greater level of clarity can be achieved.

A case study will be presented for a small data set, in which the information will be reviewed both with and without the addition of PCA. The process, interpretation and reasoning behind the use of PCA will be displayed, and will show how it can help to draw further conclusions and/or heighten confidence in initial conclusions.

40. Sparse Principal Component Analysis for Clinical Variable Selection in Longitudinal Data

Orlando Doehring, Gareth James (PHASTAR)

Background: Data collection is a time-consuming and expensive process. To minimise costs and reduce time, statistical methods can be applied to determine which variables are required for a clinical trial. Principal component analysis (PCA) is a popular exploratory technique to select a subset of variables at one timepoint. For multiple timepoints, typically each variables' measurements are aggregated, which ignores temporal relationships. An alternative method is Sparse Grouped Principal Component Analysis (SGPCA), which also incorporates the temporal relationship of each variable. SGPCA is based on ideas related to Least Absolute Shrinkage and Selection Operator (LASSO), a regularised regression technique, with grouped variables. SGPCA selects a sparse linear combination of temporal variables where each patient is represented as short multivariate time series which are modelled as a continuous smooth function of time using functional data analysis (FDA).

Aim: Compare the ability of the PCA and SGPCA to identify required variables for clinical trials.

Methods: PCA and SGPCA will be applied to a longitudinal clinical dataset to select required variables. We will compare the required variables, and the amount of variability retained for each technique under the SGPCA model.

Conclusion: This research will provide awareness of techniques to identify required variables in clinical trials, and aims to demonstrate the potential benefit of incorporating the temporal relationships in variable selection.

41. A Comparison of Analysis Procedures for Correlated Binary Data in Dedicated Multi-rater Imaging Trials

Michael Kunz (Bayer Pharma AG)

Three analysis procedures for repeated correlated binary data with no a priori ordering of the measurements are described and investigated. Examples for correlated binary data could be the binary assessments of subjects obtained by several raters in the framework of a clinical trial. This topic is especially of relevance when success criteria have to be defined for dedicated imaging trials involving several raters conducted for regulatory purposes. First, an analytical result on the expectation of the 'Majority rater' is presented when only the marginal distributions of the single raters are given. A simulation study is provided where all three analysis procedures are compared for a particular setting. It turns out that in many cases, 'Average rater' is associated with a gain in power. Settings were identified where 'Majority significant' has favorable properties. 'Majority rater' is in many cases difficult to interpret.

42. Path to Approval of osimertinib (AZD9291) in the US

Nicola Schmitt, Helen Mann, Rachael Lawrance, Laura Brooks, Alison Templeton (AstraZeneca)

The average development time of an Oncology Drug is 8.5 years. In Nov 2015, 2 years and 8 months from start of Phase I, osimertinib (AZD9291) was approved by the FDA for the treatment of patients with metastatic EGFR T790M mutation-positive NSCLC, as detected by an FDA approved test, who have progressed on or after EGFR-TKI therapy.

The choice of study design (single arm or randomized) and endpoints (objective response rate or progression free survival with independent central review versus site assessment) were critical decisions for the speed to approval, as was the degree of duration of response and safety data submitted in the NDA file. The statisticians played a pivotal role in working closely with the regulatory agency in order to initially agree a study design and an acceptable data package that would demonstrate the benefit:risk profile of osimertinib and also during the regulatory review of the data. Furthermore, the parallel development and submission of a companion diaganostic and the needs of payers and agencies worldwide had to be considered from the beginning of the development programme.

From the statistician's perspective, this presentation will share how traditional timelines were accelerated; the experience the statisticians had with the regulatory authority and how the team worked with CROs, a diagnostic partner and other providers to deliver a data package that supported approval of both the drug and companion diagnostic.

43. Overview of the Advantages and Pitfalls of Ordinal Logistic Regression

Aimie Nunn and Sue McKendrick (Quanticate)

Recovery from acute ischaemic stroke is commonly assessed using a scale of ordered categories indicating the degree of physical disability and dependence at six months following the stroke event. The majority of previous trials have used binary analyses, setting a threshold score for a favourable outcome, however, the OAST collaboration in 2007 showed that binary analyses were sub-optimal regardless of the threshold chosen, and that ordinal analyses conferred greater statistical power to detect a significant treatment effect. Despite these recommendations, uptake of ordinal analyses has been low.

We evaluated an ordinal analysis approach using data from the International Stroke Trial (IST), a large randomized factorial trial in 19,285 individuals to assess aspirin and heparin for the treatment of stroke. The functional outcome was assessed at six months using an abbreviated scale of four categories indicating full recovery (1), partial recovery (2), dependency (3) and death (4). We first present an overview of the advantages and pitfalls of ordinal logistic regression using the IST dataset. We make recommendations for analysing and reporting ordinal data where there may be doubt in the strength of the proportional odds assumption.

44. Cross-over to High Quality Graphics: Using SGRENDER to Enhance Graphics in a Cross-over Study

Martin Clancy, Simon Clancy and Sue McKendrick (Quanticate)

Good graphical summaries can enhance a clinical study report, enabling conclusions to be clearly illustrated in a concise manner. In previous versions of SAS, sophisticated graphics could be achieved using the GREPLAY procedure and ANNOTATE facilities but these were difficult to master. Summarising data graphically using the SG graphics procedures such as SGPLOT, SGPANEL, SGSCATTER is an efficient way to produce high quality graphics in SAS Version 9.2 or higher. However, some types of customisation can be difficult using SG graphics.

This poster will take a typical pharmacokinetics cross-over study as an example to show how PROC TEMPLATE can be utilised to build a template infrastructure for graphical outputs study-wide. Different studies could then use your SAS graphical templates via PROC SGRENDER. We will explain how robust code can be produced with "dynamic variables" so that high quality and consistent figures can be produced efficiently across studies.

45. Consistency of Subgroup Effects in Clinical Trials

Ring A, Day S, Schall R (Medac GmbH)

Prospectively planned confirmatory clinical trials typically investigate a single primary treatment effect but should also plan for additional exploratory analyses, such as investigating differential treatment effects in clinically relevant subgroups (e.g. gender, age or comorbidities).

While subgroup analyses conventionally are performed using a test of heterogeneity, such tests have many weaknesses and we propose to reverse the testing procedure by analysing consistency of subgroup effects using an equivalence test. For two subgroups, the scaled consistency ratio $cr=(\mu 1-\mu 2)/\sigma r$ (difference between the subgroup treatment effects, divided by the SD of the residuals) is used as a test statistic, similar to [1] for scaled average bioequivalence and [2] for carryover negligibility. An advantage of this approach is that it can be imbedded directly within a multiple testing procedure.

The analysis can be extended to more than two subgroups. In this case, for each subgroup a consistency analysis is performed by investigating the difference of the effect in that subgroup and its complement, hence analysing it as in the case of two subgroups (but adjusted for multiple testing). The presentation will discuss statistical properties of this approach and potential for implementation within clinical trials.

[1] Tothfalusi L, Endrenyi L. Limits for the scaled average bioequivalence of highly variable drugs and drug products. Pharm Res. 2003;20(3):382-9.

[2] Ocana J, Sanchez MP, Carrasco JL. Carryover negligibility and relevance in bioequivalence studies. Pharm Stat 2015, 14(5): 400-408.

46. Fraudulent Data Checks and How to Develop Them

Shafi Chowdhury, Aminul Islam (Shafi Consultancy Limited)

Fraudulent checks are now required by the FDA as part of routine data checks that must be performed. However, there can be many different methods used on many different types of data. This poster looks at some of the statistical techniques used to try and identify fraudulent data. It looks at the different types of data available, and which methods can be used on each type.

The results can also be presented in many different plots, from charts and plots to complex analysis tables. This paper will also review the different methods to see if one method gives more clear results and interpretation than others. As this is something that must be done for each trial, the paper will offer valuable suggestions as to how these checks can be performed and standardised to minimise efforts.

47. How to Improve upon Time to Statistics Report

A.De Castro MSc, J.Sauser MSc, L.Gosoniu PhD, R.Venkatesan MSc (Nestlé: Nestec S.A.)

A clinical study report (CSR) is the final milestone in any clinical trial. The interpretation of statistical results plays a key role in the CSR finalization. Often Statistics interpretation is written as a standalone document and then merged with CSR for a complete interpretation of the study results. Typically, post Database Lock (DBL), pharma companies first target the topline/first information report for the primary and key secondary results and then move on to the complete Statistics Report having interpretation for all the analyses proposed in the Statistical Analysis Plan (SAP). According to MetricsChampion.org, CMR and KMR metrics data, it takes 5-10 days for pharma companies to produce the topline/first information report post database lock and approximately 6-8 Weeks for full report. All pharmaceutical companies strive hard to reduce this time. We discuss literate programming approach to reduce this time significantly.

The term Literate programming was coined by Dr. Donald Knuth in his research paper entitled Literate programming, in the Computer Journal, 1984, 27, 97-111. Dr. Donald Ervin Knuth is an American computer scientist, mathematician, and Professor Emeritus at Stanford University.

48. Findings from a review of treatment switch adjustment methods used in NICE technology appraisals

Claire Watkins (Clarostat Consulting Limited), Jason Wang (Celgene)

Treatment switching occurs in a clinical trial when control arm patients switch to experimental therapy during the study. This often happens in oncology trials where patients switch following disease progression, and can reduce the observed survival difference. An estimate of the survival effect without switching may be of interest and several methods have been developed to estimate this. These are commonly used in Health Technology Assessment (HTA). In 2014, the Decision Support Unit of the UK HTA body NICE published Technical Support Document (TSD) 16 to provide guidance on this.

A review has been conducted of NICE technology appraisals in advanced/metastatic oncology where treatment switch adjustment has been applied. Up to September 2015, 9 TAs involving 10 trials were identified as suitable for the review, with 162-750 patients and switch rates of 7%-84%. The review will be updated for the conference. Patterns in terms of disease area, study designs, adjustment methods used, and eventual recommendations for reimbursement will be summarised. The results from different methods such as ITT, censoring switchers, excluding switchers, rank preserving structural failure time models (RPSFTM), Inverse Probability of Censoring Weighting (IPCW), and use of an external control arm will be compared. Implications and guidance for statisticians applying these methods in future trials will be summarised.

2016 PSI Conference

The Pullman Hotel, Berlin **22 - 25 May 2016**

For general conference queries please contact the conference organisers at the following address:

PSI Conference Organising Team
Durford Mill
Petersfield
Hampshire
GU31 5AZ

Tel: +44 (0) 1730 715 237
Email: PSI2016conference@mci-group.com

@PSIupdate

PSI (Statisticians in the Pharmaceutical Industry) Ltd

www.psiweb.org