MBSW 2013 Speaker Titles and Abstracts
Sharing Statistical Solutions

Version 6 May 2013

Table of Contents (click on page numbers to jump to abstract)

Monday afternoon, May 20 2013

Short Courses

Using Multiple Outcomes in Survival Models
Terry Therneau, Mayo Clinic

Survival analysis methods have played a central role in the reporting and understanding of acute disease trials. Most often they are applied to a single primary endpoint. In the last several years these methods been extended to deal with general multi-state models. These include multiple repeating events (infection, hospital admission), multiple overlapping events such as toxicities, and competing outcomes. What are the strengths and limitations with respect to what you can learn and present from a study using such endpoints, as compared to standard methods? Should you consider including these as primary or secondary outcome in a study?

This short course will focus on practical aspects of data setup and model fits, illustrated in examples of these analyses from my own experience. The primary tools are marginal and mixed effects Cox models along with simple and competing risks survival curves. One subtheme are the various pitfalls that occur: thinking about and handling time dependent covariates comprises a substantial portion of the latter issue.

Analysis Methods for Informative Censoring
Diane Fairclough, University of Colorado Denver

Non-ignorable missing data is not uncommon in longitudinal studies where participants may experience morbidity or mortality. This workshop will examine a number of models that are commonly proposed for the analysis of these studies. The models include pattern mixture models, joint mixed-effects and time-to-effect model, multiple imputation using MCMC algorithms and last observation carried forward.. Specifically, we will examine both the underlying assumptions and the practical constraints these classes of models. The procedures for implementing each of the models will be presented and illustrated using data from a clinical trial measuring HRQOL in patients with advanced lung cancer.

After completing this workshop, participants will be able to:

1. Describe the assumptions underlying each of the models.
2. Identify which methods can be used in a specific study given characteristics such as equal vs. unequal interval assessments, monotone vs. non-monotone missing data, linear and non-linear change patterns.
3. Perform sensitivity analyses using these models where appropriate.

Participants should have some experience with mixed-effects (or hierarchical) models.

Plenary Session

Information Allergy

Frank Harrell, Vanderbilt University

Information allergy is defined as (1) refusing to obtain key information needed to make a sound decision, or (2) ignoring important available information. The latter problem is epidemic in biomedical and epidemiologic research and in clinical practice. Examples include

- ignoring some of the information in confounding variables that would explain away the effect of characteristics such as dietary habits
- ignoring probabilities and “gray zones” in genomics and proteomics research, making arbitrary classifications of patients in such a way that leads to poor validation of gene and protein patterns
- failure to grasp probabilistic diagnosis and patient-specific costs of incorrect decisions, thus making arbitrary diagnoses and placing the analyst in the role of the bedside decision maker
- classifying patient risk factors and biomarkers into arbitrary “high/low” groups, ignoring the full spectrum of values
- touting the prognostic value of a new biomarker, ignoring basic clinical information that may be even more predictive
- using weak and somewhat arbitrary clinical staging systems resulting from a fear of continuous measurements
- ignoring patient spectrum in estimating the benefit of a treatment

Examples of such problems will be discussed, concluding with an examination of how information–phobic cardiac arrhythmia research contributed to the deaths of thousands of patients.

Expert Systems and Predictions for the Future

Lee Wilkinson, Skytree, University of Illinois at Chicago

Abstract:
Statistical computing in the next decade will be influenced more by technology than by developments in statistics. This forecast is based on contemporary observations of the field during the last 40 years and on a supposition that extrapolating these trends is not unreasonable. The technology driving this forecast includes not only hardware, but also the software that provides the infrastructure for individual and community interaction with computers. We should not be surprised to see a proliferation of intelligent data analysis systems embedded in everyday objects and Web sites; automated visualizations for data discovery; analytic systems that are accessible by non-statisticians (a trend toward simplicity and away from comprehensiveness); distributed analytic systems that talk to each other, fuse disparate data in real time, and draw conclusions on the evidence; and communities of open-source developers exceeding the scope and capabilities of commercial companies. Whether computer scientists eventually take over this field will depend on how actively statisticians participate. Statisticians interested in statistical computing and its future incarnations will have to engage in joint research with computer scientists in order to continue to have an influence.

Clinical Track

Safety in Drug Development: A Panoramic View of New Solutions and Statistical Approaches

Meta-Analysis of Clinical Trial Safety Data in a Drug Development Program: Answers to Frequently Asked Questions

Brenda Crowe, Eli Lilly

Meta-analyses of clinical trial safety data have risen in importance beyond regulatory submissions. During drug development, sponsors need to recognize safety signals early and adjust the development program accordingly, so as to facilitate the assessment of causality. This talk will focus on common questions encountered when designing and performing a meta-analysis of clinical trial safety data. Though far from an exhaustive set of questions, they touch on some basic and often misunderstood features of conducting such meta-analyses. The questions address the following topics: choice of studies to pool, effects of the method of ascertainment, use of patient level data compared to trial level data, the need (or not) for multiplicity adjustments, heterogeneity of effects and sources of it, and choice of fixed effects versus random effects.

Simulation-Based Designs for Safety Monitoring in Clinical Trials: Two Case Studies

Fei Chen, Johnson and Johnson

Regulatory agencies have shown increasing concern about potential, specific (usually rare) safety signals during clinical drug development. Entire classes of drugs have been placed on clinical hold because of such concerns, with regulators requiring safety monitoring rules for reducing patient risk in ongoing development programs as a pre-condition for lifting the clinical hold. Multiplicity issues arise naturally in this context, both from the repeated looks into accumulating, unblinded data, but also due to the various studies included in a program. From a regulatory perspective, the associated false positive rate is less of a concern, but from the sponsor point-of-view it can make investing in the development program unfeasible. This talk will describe the experience of developing a safety alert rule in a late phase clinical development program, presenting different alternatives considered and their respective operating characteristics at the trial and program levels. More generally, the issue of safety signal detection vs. demonstrating lack of a clinically relevant safety signal will be discussed in the context of safety monitoring strategies to achieve a balance between protecting patients and keeping drug development economically viable.

Evaluating Change in Hazard in Clinical Trials With Time-to-Event Safety Endpoints

Rafia Bhore, Novartis

In drug development, evaluation of safety of a new treatment is often made using long-term safety follow-up data (whether controlled or uncontrolled). Signals of adverse events can be then assessed through time-to-event analysis of safety endpoints that are based on either clinical events or markers signaling serious adverse events that are not directly observable. After certain duration of exposure to the drug, it becomes important to know whether the risk of an adverse event remains constant, decreases or increases over time. Escalation of risk is important to detect as quickly as possible, particularly when risk is constant over a short period of time and suddenly takes a wrong turn, that is, starts to escalate over time. This is defined as a statistical problem on whether the hazard changes over time.

In this talk, we present change point methodology for estimation and testing of changes in hazard (function) for time-to-event endpoints. We illustrate the use of piecewise exponential model with piecewise constant hazard to model the hazard function. Using likelihood-ratio test and the corresponding likelihood-ratio based approximate confidence regions for the unknown change point (Loader 1991) we show how to test for change in hazard and estimate the change point. Furthermore we discuss bootstrapping as an alternative to estimate the time of change in hazard for the adverse event of interest.

Recommendations on Summaries of Integrated Phase 1 Safety data

Sveta Weiner, Johnson and Johnson

Over the years, most companies are generating integrated safety summaries of adverse events, laboratory measurements, vital signs and ECG measurements for Phase 1 studies (early development and Clinical Pharmacology studies). However, regulatory guidance is limited on this topic, which leads to different methods of analysis. Some of the challenges in pooling large number of phase 1 studies are due to varying study designs (parallel group, crossover, single-sequence), populations (healthy, patients, special population such renally impaired, elderly), dosing (single dose vs. multiple doses) and concomitant medications (drug-drug interactions studies) and lack of placebo in most studies. With such a vast scenario, we will evaluate pooling strategies and their pros and cons. Recommendations on reasons for pooling, methods of analysis and impact on the interpretation of results will be made.

Discovery/Preclinical Track

Current Statistical Issues in Assay Development and Validation

Directly Testing the Linearity Assumption for Assay Validation

Steven Novick, GlaxoSmithKline and Harry Yang, MedImmune

The ICH Q2(R1) guideline for testing linearity in validation of analytical procedures suggests that “linearity should be evaluated by visual inspection of a plot of signals as a function of analyte concentration or content.” The EP6-A guideline recommends more quantitative methods that compare straight-line and higher-order polynomial curve fits. In this talk, a new equivalence test is proposed to compare the quality of a straight-line fit to that of a higher-order polynomial. By using orthogonal polynomials and generalized pivotal quantity analysis, one may estimate the probability of equivalence between a straight line and a polynomial curve fit either in the assay signal space (the ”Y” values) or in the concentration space (the “X” values). In the special case of the linear-to-quadratic polynomial comparison, an equivalence test may be constructed via a two one-sided T test.

Development and Validation in the World of Quality by Design

Tim Schofield, MedImmune

The implementation of the USP Bioassay chapters as well as other guidelines on analytical method development and validation requires consensus on the goals and statistical approaches associated with various elements of these activities. There is common confusion regarding the definition of linearity and its assessment, while there exists an array of limits and approaches for defining these limits in the context of the use of a method. Lack of understanding of equivalence testing and a Quality by Design approach to the method lifecycle interferes with achieving progress towards a risk based approach to method development and analysis. This talk will highlight the practical and regulatory issues related to these fundamental elements of analytical methods, and offers suggestions to the practicing statistician who feels they are lost behind the looking glass.

Statistical Issues in Development and Validation of Functional Assays

Tsai-Lien Lin, FDA

Functional assays are often preferred by clinical and biological scientists to measure the immune responses to vaccines, which are used as surrogate endpoints in clinical studies of vaccine efficacy, because what functional assays measure generally correlate better with protection. However, functional assays pose greater statistical challenges during assay development and validation. In addition to the much larger inherent variability involved with functional assays, there are many other problems that could potentially make the assays unfit for intended uses. In this presentation, several statistical issues and challenges recently encountered in a few functional assays will be discussed and illustrated with examples.

Observational Track

Econometric Methods in Pharmaceutical Research

Discrete Choice Methods applied to Prescribing Decisions

Mariana Carrera, Case Western Reserve University

Economists use discrete choice models to study how decision makers choose between alternatives. It is a natural class of models for analyzing prescribing decisions within a class of drugs. This session will provide a basic overview of these methods, with particular attention paid to the multinomial logit model. We will discuss an application of this model to the study of how statin prescribing responds to variation in patient copayments.

Limited Dependent Variables in Health Research: Some Examples From Smoking

Erik Nesson, Ball State University

This paper estimates the effects of tobacco control policies on non-smoking workers' exposure to secondhand smoke at their jobs. I use a novel measure of workers' self-reported exposure to secondhand smoke at their jobs from the National Health and Nutritional Examination Surveys and combine this self-reported measure with a biomarker of individuals' recent nicotine exposure to test whether any decrease in self-reported secondhand smoke exposure translates to reduced overall nicotine exposure. While I find little evidence that cigarette excise taxes or prices reduce workers exposure to secondhand smoke, I find evidence that workplace and restaurant or bar smoke-free air laws reduce secondhand smoke exposure and these reductions translate into reduced overall nicotine exposure. I more directly test whether workplace and restaurant or bar smoke-free air laws reduce overall secondhand smoke exposure through changes in work exposure by estimating specifications which interact the level of reported workplace exposure with tobacco control policies. I find some evidence that this reduction in nicotine exposure comes from reductions in secondhand exposure at work and evidence that smoke-free air laws reduce secondhand smoke exposure through other pathways as well.

Using Claims Data to Analyze Adherence to Pharmaceutical Therapies: Initial Treatment and Time to Discontinuation

John Bowblis, Miami University

The use of secondary claims datasets to study the use of pharmacological interventions has become increasing popular with the use of electronic medical records. However, because claims data is not generated in a controlled, experimental setting, there are multiple econometric issues that need to be addressed in order to not obtain biased results. This presentation will examine these econometric issues in the use of secondary datasets when related to HEDIS quality measures related to initial treatment and time to discontinuation of pharmacological therapy. Drawing from these examples, we will investigate sample selection issues that may arise when claims data is drawn from various sources that have different insurance plan designs. Additionally, we will discuss issues related to the endogeneity of treatment choice and how to address endogeneity with instrumental variables techniques, including a discussion of the exclusion restrictions, exclusion restrictions validity tests, and two-stage residual inclusion.

Manufacturing Track

Stability Analyses: Beyond ICH Q1E

Statistical Considerations for Mitigating the Risk of Individual OOS Results on Stability

Jeff Gardner, DataPharm Statistical and Data Management Services

Mixed-effects modeling of stability data (where lot intercepts, lot slopes, or both are treated as random effects) offer superior methods of estimating product expiry to those described in ICH Q1E. Whether basing expiry on the worst-case observed batch slope (per Q1E) or basing expiry on the distribution of batch slopes (e.g. mixed-effects modeling), there is still risk of observing individual out-of-specification (OOS) results over the course of the established product shelf-life. The occurrence of such results can have significant quality and regulatory impact, as a manufacturer must then issue a field alert and consider recalling product and/or reducing the product's expiry period. This presentation examines the risk of future batches producing individual OOS results when using the above approaches and discusses ideas for mitigating this risk.

On the Shelf-Life of Pharmaceutical Products

Jim Schwenke, PQRI, Applied Research Consultants

This presentation is a discussion on the recent article published by the Product Quality Research Institute (PQRI) Stability Shelf Life Working Group (Capen, R.C., et.al., 2012. "On the Shelf Life of Pharmaceutical Products", AAPS PharmSciTech, Vol. 13, Issue 3, 911-918). New terminology is proposed that distinguishes between different shelf life concepts involved in the discussion of the stability of pharmaceutical products. Such comprehensive and common language is currently lacking from various guidelines, which confuses implementation and impedes comparisons of different methodologies. Five terms being proposed allow for a coherent discussion of shelf life: true shelf life, estimated shelf life, supported shelf life, maximum shelf life, and labeled shelf life. These

concepts are already in use, but not named as such. Various levels of “product” on which different stakeholders tend to focus (e.g., a single dosage unit, a batch, a production process, etc.) are discussed with respect to the proposed terminology. A key missing element in the discussion of shelf life is a Quality Statement, which defines the quality standard for all key stakeholders. The presentation concludes with comments on the current ICH methodology for estimating shelf life. Some preliminary comments on the Working Group's proposed estimation methodology are presented.

Change During Patient Use - Questions and Challenges

Rebecca Elliott, Senior Research Scientist

According to ICH Q1E, shelf-life and expiry can be justified by looking at long term stability changes. However, there are other potential changes to the product. Some refrigerated multi-use products and reconstituted products have a period of time for patient or medical professional use, a time when the product does not have to be refrigerated. The change over this period may be different than that of the long term stability. When products are shipped, they may experience higher temperatures and humidities than the long term stability conditions. How are these potential changes incorporated into the justification for specifications and dating? What analysis challenges may arise from these data?

Tuesday afternoon, May 21 2013

Clinical Track

Quantitative Methods for Decision Making

The Use of Global Benefit-Risk Assessment in Drug Development

Yili Pritchett, Astellas

Data from randomized clinical trials are typically analyzed by efficacy and safety separately. However, understanding the overall benefit and risk profile of a treatment and making comparisons between therapies from benefit-risk point of view are often needed. Chuang-Stein et al (1991) proposed to construct a consolidated measure, namely Global Benefit-Risk (GBR) assessment, to conduct such evaluations. In this talk, applications of GBR method in various type of clinical trials from studies with clear defined efficacy measures to those where outcomes are all detrimental events will be reviewed, statistical considerations with respect to sensitivity and robustness of the method will be presented, and the role of this approach in drug development will be discussed.

Evaluation of Benefit-Risk in Clinical Programs

Freda Cooner, FDA

As the medical products become more complicated and personalized medical treatment attracts more attention, there is a growing demand on benefit-risk assessment research. Many different approaches for the quantitative and qualitative evaluation of benefit-risk of medical products have been studied extensively. However, great difficulties of developing a consistent benefit-risk assessment arise from the fact that each medical product may demand different efficacy and safety endpoints, with varying timescales and degree of impact on patients. We, within this presentation, will present and evaluate several systematic benefit-risk assessment techniques that could potentially be adopted as a universal approach. Details of some models will be discussed; and some Advisory Committee meeting examples on benefit-risk evaluation will be provided.

Benefit-and-Risk Evaluation of Treatment of Arterial Disease in the Lower Limbs

Justin Recknor, Gore

There are many different treatments available for stenosis or occlusion in the superficial femoral artery (SFA). Three of the treatment options which are commonly used are percutaneous transluminal angioplasty (PTA), stenting, and surgical bypass. Assessment of both the risks and benefits of each of these treatments can be quite difficult under the constraints encountered with each of these interventional procedures. In depth descriptions of each of the procedures, methods of evaluation, and issues encountered will be discussed.

Predicting Phase III Success Using a Combination of Phase II Results and Historical Reference Data

David Burt, GlaxoSmithKline

Traditionally, the decision to progress a pharmaceutical asset to late stage development has been based on the results of early, relatively small Phase II clinical studies. Positive results in these early studies are often not reproducible in the later studies. In our presentation, we will explore a method of predicting the reproducibility of the Phase II results that uses these results in combination with meta-analysis parameter estimates obtained from historical study placebo response rates. A case study example will be discussed and general performance properties will be evaluated via a simulation study.

Discovery/Preclinical Track

Next Generation Sequencing and Its Statistical Issues

Next-Generation Sequencing - Opportunities and Challenges

Phil Ebert, Eli Lilly
John Calley and Thomas Barber, Eli Lilly

Next-generation sequencing has revolutionized nucleic acid research, dramatically increasing the rate at which we acquire new data as well as the amount of data being generated. Utilization of this technology for interrogation of the transcriptome has led to new insights into the expression and regulation of RNA, extending beyond traditional differential expression analysis to digital quantitation at the single transcript level. As with any new technology, there are growing pains, including the methodologies utilized to assess the quality of the data and quantify the significance of results. For the full potential of RNA sequencing to be realized these challenges need to be addressed. A crucial component of this will be developing novel statistical approaches to evaluate the impact of data quality as well as sampling uncertainty on predictions of differential expression at both the gene and isoform level.

Statistical Considerations for NGS Study

Ray Liu, Millennium

Genome-wide association studies (GWAS) utilizing common single nucleotide polymorphisms (SNPs) have found thousands of common genetic variants associated with a wide variety of diseases and other complex traits. However, for most complex traits, heritability explained by identified SNPs is low. For example, variants that are discovered and validated by GWAS account for only 5-10% of variation in height. It has been hypothesized that some of the missing heritability may be due to the effect of rare variants (< 1% in population frequency) which are previously unidentified with DNA microarrays designed for GWAS. With the advances in next generation sequencing (NGS) technology, it is now possible to genotype whole genomes and explore the association between rare variants and unexplained heritability. The major analytical challenge is how to identify the association between rare variants and phenotypes, given that there is little power when testing one variant at a time. This presentation aims to provide an introduction of statistical methods for design and analysis of genome-wide sequencing association studies. Methods to be discussed include CAST, CMC, EREC, C-alpha, and SKAT.

Variation Discovery From Heterogeneous Tumor-Normal Paired DNA Samples Using Exome Sequencing

Hyonho Chun, Purdue University

Next-generation sequencing (NGS) technology has become popular in the understanding of the molecular mechanisms underlying cancer. Exome sequencing provides a cost-effective and time-efficient tool, and hence we now able to systematically explore the structural basis of cancer from multiple samples. However, in practical applications, analytical complexity frequently limits the power of NGS discovery in large-scale sequencing projects. Such complexity includes sample contamination and tumor heterogeneity. In this work, we first focus on cataloguing tumor mutations via tumor identity check and purity estimation from mutation calling algorithm across multiple samples, and then perform tumor-subtype classification based on the identified mutation signatures. Simulations and real data analysis will be conducted to show the advantages of the proposed method.

Observational Track

FDA Sentinel/Mini-Sentinel and OMOP Initiatives: Status Update

FDA Sentinel/Mini-Sentinel: Developing a National Active Surveillance System
Azadeh Shoaibi, FDA

This presentation will cover a variety of important topics regarding FDA's efforts in the area of active medical product surveillance. The topics include the following:

- An update on FDA's Sentinel Initiative, including current activities and future plans
- An overview of the Mini-Sentinel pilot and its accomplishments
- An update on current and future collaborations and initiative expansions

OMOP Overview and Scientific Results
Bill DuMouchel, Oracle Health Sciences

In partnership with PhRMA and the FDA, the Foundation for the National Institutes of Health launched the Observational Medical Outcomes Project (OMOP, see http://omop.fnih.org), a public-private partnership. This interdisciplinary research group has tackled a surprisingly difficult task that is critical to the research community's broader aims: identifying the most reliable methods for analyzing huge volumes of data drawn from heterogeneous sources, with emphasis on discovering and confirming associations involving adverse drug reactions. OMOP has conducted a series of experiments to generate empirical evidence about the performance of observational analysis methods in their ability to identify true risks of medical products and discriminate from false findings. Results and conclusions so far from these experiments will be described.

High-Dimensional Propensity Score: Lessons Learned
Jeremy Rassen, Brigham and Women's Hospital and Harvard Medical School

The High-Dimensional Propensity Score (HDPS) algorithm is a very prominent and upcoming methodological approach to adjust for confounding factors in large administrative health care databases. This presentation will share some key findings and provide recommendations on how to optimally utilize HDPS.

Evolution of Active Surveillance: Industry Perspective
Stephen Motsko and Kenneth Hornbuckle, Eli Lilly

This session will further complement the other presentations within this session by providing an overview of active surveillance initiatives in other geographical regions. This presentation will also provide an overview of areas to consider to implement an active surveillance strategy as part of a comprehensive post-marketing surveillance system, including how are several PhRMA companies approaching active surveillance, scope and definitions for active surveillance, describe key features and challenges of active surveillance, provide insights on how active surveillance can support post-marketing activities and supplement the decision making process. Lastly, the presentation will share a case study of an active surveillance plan for a specific marketed product.

Manufacturing Track

Accelerated Testing for Biologics

Physical Biochemistry, Metrology, and Accelerated Degradation Experiments

William R. Porter, Peak Process Performance Partners

The design of accelerated degradation experiments for biologics requires an understanding of the complex multifactorial sources of variation in chemical and physical factors that can cause degradation. Biologics degrade by a multitude of chemical mechanisms involving hydrolysis, oxidation, photolysis and racemization. An understanding of the limited precision of many analytical methods used to characterize biologics is also critical to experimental design. Designs which favor measurement of time to failure (a knowable quantity) are superior to designs which rely on estimation of reaction rates (which require understanding of mechanisms and often place demands on assay precision that are difficult to meet). Arrhenius models and cycling experimental designs can address these various factors affecting stability.

Accelerated Stability Modeling for Bioproducts

Kevin Guo, Eli Lilly
Adam Rauk, inVentiv, William Weiss, Suntara Cahya, Eli Lilly

In developing bioproducts, a short-term multiple temperature accelerated stability study is performed to expedite the selection of a formulation with an acceptable stability profile throughout a shelf-life of at least 24 months. Based on this accelerated stability study, a prediction model is developed to estimate the long-term stability profile of the formulation candidates under the intended long-term storage condition. In this presentation, design considerations for accelerated stability studies, predicting long-term stability using the Arrhenius equation, and the direction of future work will be discussed.

Modeling Sub-Visible Particle Data With Product Held at Accelerated Stability Conditions

Jose Ramirez, Amgen

For performance parameters that can be approximated by a normal distribution, drug product stability at accelerated conditions can be evaluated using a random coefficients model. Sub-visible particle data are, by nature, discrete counts that should be modeled using an appropriate distribution like the Poisson. In this talk we discuss the structure of accelerated stability data and the different components that need to be taken into account, such as correlations between different particle sizes and autocorrelation of sampling times. Via a simulated example we show how to fit a repeated measures random coefficients Poisson model and the inferences that can be made with such a model.

Determining Equivalence Acceptance Criteria With Accelerated Stability Data
Leslie Sidor, Amgen
Rick Burdick and Camilla Santos, Amgen

The use of statistical equivalence testing for providing evidence of process comparability in an accelerated stability study is advocated over the use of a test of differences. The objective of such a study is to demonstrate comparability by showing the stability profiles under non-recommended storage conditions of two processes are equivalent. Because it is difficult at accelerated conditions to find a direct link to product specifications, and hence product safety and efficacy, an equivalence acceptance criterion is proposed that is based on the statistical concept of effect size. As with all statistical tests of equivalence, it is important to collect input from appropriate subject matter experts when defining the acceptance criterion.

Student Track

How to Nurture Your Statistics Degree into a Profession in Pharma

The Drug Development Process and the Role of Statisticians
Yun-Fei Chen, Eli Lilly, Cathie Spino, U. Michigan

From Courses to Careers: A View From the Pharmaceutical Industry
Brad Evans, Pfizer, Haoda Fu, Eli Lilly, Jackie Reisner, PPD

Leadership Skills for Statisticians: Why it is Important and How to Develop Them
Gary Sullivan, Eli Lilly

Tuesday evening, May 21 2013
Leadership and Career Development
Paul McKenzie, Johnson and Johnson

Clinical Track

Contributions of Statisticians in Clinical Trials: From Design to Analysis

The Role of Statistics Innovation in Improving the Productivity of Pharma R&D

Haoda Fu, Eli Lilly

Drug development has become increasingly costly, lengthy, and risky. The call for better decision making in research and development has never been stronger. Analytic tools that utilize available data can inform decision makers of the risks and benefits of various decisions, which could lead to better and more informed decisions. In this talk, we share a few examples on using advanced analytic methods to help drug development and improve decision making.

New Insights in Use of Baseline Covariates in Clinical Trial Design and Data Analysis

Yongming Qu, Eli Lilly

Effective use of baseline covariate in statistical analysis can considerably improve the estimation efficiency. We discuss two specific issues on the use of baseline covariates in clinical trials. First, if subjects are randomized to treatments based on a stratification factor derived from a continuous baseline variable, should the continuous baseline variable or the baseline strata variable be adjusted? Second, for generalized linear models, if baseline covariates are adjusted, what is the impact on the estimation of the group means and how can it be adjusted for?

Examination of Analysis Methods for Positive Continuous Dependent Variables: Model Fit and Cost Saving Implications

Brian Smith, Amgen
Maria De Yoreo, Department of Applied Mathematics and Statistics, University of California, Santa Cruz

In most if not all fields of research variables are often both continuous and restricted to be positive. We analysed 70 data sets that were continuous and positive from clinical and pre-clinical trials. Different models (log transformation, change from baseline, etc.) were examined and the fit results were compared. On average including baseline as a covariate decreases sample size about 70% as compared to a model which ignores baseline or by 20 to 33% compared to a baseline-adjusted response model. Additionally, log-transformation appears to decrease the sample size needed on average between 20% to 35%. We draw 3 conclusion from this work. 1) If a baseline is available, use of baseline as a covariate should always be undertaken. 2) Although we recommend exploration of data from previous studies, percent change from baseline analyses should not be undertaken unless there is strong empirical evidence that for that endpoint it is preferred. 3) Again with the caveat that nothing replaces exploration of data from previous studies, log-transformation ought to be the default analysis of positive data unless exploration of previous data provides convincing evidence that the natural scale is preferred.

Detection of Outliers in Biomedical Data

Rahmatullah Imon, Ball State University

Identification of outliers has been an area of great deal of attention for many years. The detection and handling of outliers are essential because their presence could create huge interpretative problems. The problem of the detection of outliers is largely resolved for single outlier in univariate data but many problems are still open for multiple and multivariate outliers because of their masking (false negative) and swamping (false positive) effects. Identification of outliers are even more complicated in Biomedical data because it varies from one extreme to another: rare data sets whose distribution is unknown, small data sets with high dimensions, and large data sets that do not follow any known distribution. Nonlinear models are more prevalent in biomedical modeling which make the problem of outlier detection even more cumbersome. In this paper we present an overview of the major developments in the area of detection of outliers in biomedical data. Within the field of statistics, there are two general approaches to outlier identification: methods based on robust distances, and methods based on projection pursuit. Both of these two methods have wide applications in the detection of outliers and they have limitations too. The diagnostic-robust approach has become more popular now to overcome some of their limitations. Given the massive datasets that have recently become available, it is not surprising that the machine learning community has also taken an interest in outlier detection. Most of the outlier detection techniques in data mining consider principal components, distance, and density-based methods, since these are most applicable to the large datasets with high dimensions that have become more prevalent in recent years. The major algorithms within each category are briefly discussed in this paper together with current challenges and possible directions of future research.

Discovery/Preclinical Track

Reproducible Research

The Reproducibility Initiative: a potential solution to the irreproducibility problem

Elizabeth Iorns, Science Exchange

The quality of research reported in the literature is currently under intense scrutiny due to the inability of major research institutions to reproduce the majority of published research findings. Unfortunately current measures of presumed quality such as impact factor, citation rates and numbers of independent publications reporting similar results are not able to identify robust reproducible results from the literature. However, direct replication experiments were able to identify robust reproducible results that then tended to validate in multiple experimental models. Direct independent replication therefore provides an efficient screening mechanism to identify high quality reproducible research.

Independent replication also provides important validation for reagents and methods used in research studies. Part of the problem that underlies the irreproducibility of research is that reagents and methods vary in their robustness. For example, antibody suppliers differ in the amount of validation they perform and the information supplied with the reagent.

The Reproducibility Initiative Validation Service provides a platform to rapidly and cost effectively screen studies, methods and reagents via independent validation of key experimental results. This allows the identification of high quality reproducible research and reagents.

Studies, protocols or reagents can be submitted via an online portal at https://www.scienceexchange.com/validation for independent validation by expert scientific service providers from top research institutions. Submitted experiments are blindly matched with an appropriate, verified provider who then reproduces the experiments on a fee-for-service basis. All submissions are confidential, with resultant data and findings kept private. Upon completion, all experimental results will be provided to the submitter along with a certificate of reproducibility and permission to use the Reproducibility Initiative “independently validated” badge, if results are successfully validated.

The initiative takes advantage of the Science Exchange network of providers who operate on a fee-for-service basis where their only incentive is to produce high quality data, not a specific result. The network currently has more than 1000 academic core facilities and commercial service providers (CROs), and more than 1400 different types of experimental services are available from these experts. The network provides an opportunity to easily access specialized scientists to validate research.

Barriers to Reproducible Research and a Web-Based Solution
Matt Shotwell, Vanderbilt University

Reproducible research methods aim to establish a record of research activities, so that others can more easily reproduce and evaluate scientific findings. In applied statistics, reproducible research means fully documenting or scripting all data analysis plans and procedures. Despite the appeal, reproducibility is not uniformly practiced in biostatistics. A survey conducted within the Department of Biostatistics at Vanderbilt University has identified several practical barriers to adopting reproducible practices. This work addresses some pragmatic measures to overcome these barriers using a web-based reproducible research workflow and free software tools, including GNU make, the GIT revision control system, and R utilities for literate programming.

Doing Reproducible Research Unconsciously: Higher Standard But Less Work
Yihui Xie, Iowa State

Traditionally we finish statistical computing before writing reports, and the results written into the reports are essentially "dead". To update the reports, we have to redo computing. In this talk, we introduce live documents that generate reports directly, with results dynamically obtained from computer code. We show how we can change the research workflow and present results in new ways with the help of tools such as RStudio + knitr and modern technologies like HTML5 and WebSockets. Reproducible research should be so natural that we do not even notice it.

Observational Track

Methods for Missing Data

Towards a Complete Solution for Cost/Effectiveness in Oncology: Handling Heterogeneity, Variability and Censoring

Gerhardt Pohl, Eli Lilly

The Incremental Cost Effectiveness Ratio (ICER) is the primary tool of health technology assessment around the globe. The ICER is defined as the difference in mean cost between treated and comparator groups divided by the difference in mean effectiveness. To provide a complete analytic solution for determining the ICER in oncology, one would desire to address issues of variability in the estimate, heterogeneity in risk factors (such as age and tumor stage at diagnosis , etc.) and administrative censoring (patients alive at last contact). We propose a method based on bootstrapping patient costs and survival to provide an indication of variability. To address heterogeneity, patients are stratified based on prespecified baseline factors. Within each stratum, censoring of survival is handled by the Kaplan-Meier method; and censoring of costs, by the method of Lin et al., Biometrics, 1997. The within-stratum cost and survival differences are pooled proportionally to the stratum size to form overall mean cost and survival differences. These are subsequently used as the basis for calculating the ICER in each bootstrap iteration. Connections between this stratum weighting and inverse propensity score weighting are made explicit as well as connections to concepts in casual inference such as average overall treatment effect and average treatment effect among the treated.

Notes on Joint Analysis of Longitudinal and Time-to-Event Data by Random Effects Models

Lei Liu, Northwestern University

In this paper we consider three questions in joint random effects models of longitudinal and time- to-event data. First, using two real datasets, we compare two types of dependence of survival outcome on the longitudinal measures: on current underlying value or on random parameters (e.g. random intercept and random slope) separately. Second, we explicitly distinguish informative drop-out and dependent terminal event. We show that the common assumption of conditional independence on random effects between longitudinal measures and time-to-event might not hold when the event is terminal. Third, we discuss the controversy in the inclusion of individual frailty term in the survival outcome. This paper helps clarify confusions in the joint modeling of longitudinal and time-to-event data, making such models more accessible to general statistical practitioners.

Summarizing Longitudinal PRO's in the Presence of Limited Survival

Li Li and Gerhardt Pohl, Eli Lilly

In this talk we focus on a special case of informative missing data, patients who cease to provide longitudinal patient reported outcomes (PRO's) due to death. We compare various commonly used methods for analyzing such data and propose an approach that maintains the underlying ordered characteristics of the response without assuming the scale to be interval in nature. One commonly used method is mixed-model repeated measures (MMRM) with different mean outcomes at each time point. The assumption

underlying this method is that patients have the same response over time with any variability among patients appearing only in the PRO dimension. In reality, patients may have variable time scales with little sense of a shared common response at the similar time points. Other simple approaches such as calculating area under of the PRO outcome curve over time also fail to provide an effective solution. This approach assumes time and PRO level to be interchangeable. The approach we explore is to compare the time spent in the various PRO states supplemented with the death as a worst category of the PRO. Descriptive statistics and hypothesis testing using this approach are discussed.

Manufacturing Track

Spec Setting: What Does it Mean/What Goes Into it?

Drug Product Specifications: Considerations for Developing Product Quality Demonstration Methods

Helen Strickland, GlaxoSmithKline

According to ICH Q6 drug product "specifications are critical quality standards that are proposed and justified by the manufacturer and approved by regulatory authorities as conditions of approval". For most drug products, these critical quality standards are typically registered in terms of an analytical test method (in vitro test), a sample size, and acceptance criteria applied to each test result. If the test results meet the acceptance criteria then the tested batch can be viewed as having demonstrated suitability for patient use in respect to that "physical, chemical, biological, or microbiological property or characteristic" (ICH Q8(R)) tested. From a statistical perspective, this resembles the basis of a statistical hypothesis test. However, in most cases, the hypothesis has not been explicitly defined and the Type I and Type II error rates have not been quantified. The purpose of this session is to explore the concept of defining a drug product specification in terms of a population requirement and then constructing a demonstration test based on well defined statistical properties such that the probability of making an incorrect decision regarding the suitability of a drug product for patient use is quantified.

Clinical Considerations for Developing Product Quality Requirements for Uniformity of Dosage Units

Tim Kramer, Eli Lilly

Relating Dose Response Curves and Patient Variability to Content Uniformity
Pharmaceutical companies frequently conduct studies to understand the effects of drug product processing parameters on content uniformity. Typically, content uniformity is evaluated directly against USP criteria or, indirectly, by comparing results to criteria derived using Bergum’s method. In a time when the pharmaceutical industry and regulators are increasingly focusing on the impact to patients when setting specifications and process requirements, the criteria for content uniformity seem more stringent than required. To quantify the in-vivo impact of content uniformity, the effect of varying degrees of content uniformity on the variability in the Cmax and AUC for both individual and groups of patients in five dose-response studies were evaluated. This talk will summarize the results and implications of that analysis for content uniformity requirements. In addition, some content uniformity criteria that augment patient-centric requirements will be presented.

Challenges in Process Comparison Studies

Seth Clark, Merck

Biological products such as mAbs have complex bioprocesses to derive, purify, and formulate “drug substance” (DS) and “drug product” (DP). As process improvements are made during development, changes to the process that impact critical quality attributes related to safety and efficacy of the DS or DP need to be evaluated to ensure that there are no adverse impacts on quality and that previous in-vivo (preclinical and clinical) assessments remain valid after the process change. There are a number of practical issues conducting such an evaluation, the foremost being the “sample size” necessary to draw meaningful conclusions as process knowledge evolves that protects consumer risk as first priority and producer risk as second priority. This talk will discuss challenges in process comparison study design and analysis and offer some insight and solutions to some of the issues.