Volume 27, No. 1 • Fall 2019/Winter 2020

BIOPHARMACEUTICAL REPORT

CONTENTS

Note from the editors

Welcome to the first issue of the Biopharmaceutical (BIOP) Report for 2020 (or the last for 2019)! This issue's featured article, by **Wei Wang, Mary Nilsson, Rebeka Revis,** and **Brenda Crowe** (Eli Lilly), presents a succinct and insightful view on interactive visualizations for reviewing clinical safety data. A paper by **Bob Obenchain** (Risk-Benefit Statistics LLC and an esteemed Lilly retiree) presents his methodology for analyzing observational data. The methodology is termed "NU learning" (Nonparametric and Unsupervised) and is also known by many as "Local Control". He introduces his R package LocalControlStrategy. A brief paper by **Sergei Leonov** (CSL Behring) presents an insightful summary of the use of Quantum Computing in statistics.

In anticipation of the upcoming 40th anniversary of the ASA Biopharmaceutical Section (in 2020) we continue a series of vignettes from some of the key contributors to the Biopharma who reflect on the past and offer insights for the future. This issue contains vignettes by **Craig Mallinckrodt** (Biogen) and **Steven Snapinn** (Independent Consultant).

Gary Sullivan (Espirer Consulting) presents an insightful note on the nature of statistical leadership. **Ilya Lipkovich** (Eli Lilly) and **Alex Dmitrienko** (Mediana) presents a qualitative summary of a recent BIOP survey among the editors of several applied statistical journals in Pharma on DO's and DON'Ts when submitting a statistical manuscript.

This issue also presents updates from the Nonclinical Biostatistics Working Group by **Steven Novick** (AstraZeneca) and **Xin Huang** (Abbvie).

The issue is concluded by a presentation of the upcoming book, "Real World Health Care Data Analysis: Causal Methods and Implementation Using SAS", by **Douglas Faries, Xiang Zhang, Zbigniew Kadziola, Uwe Siebert, Felicitas Kuehne, Josep Maria Haro,** and **Robert L Obenchain**.

We would like to take this opportunity to thank **Ilya Lipkovich** for his service as the BR Editor in 2019 and welcome the new members of the editorial board: **Xiaofei Wang** (Editor) and **Peter Mesenbrink** (Associate Editor).

We hope you enjoy reading this issue and welcome feedback, suggestions for improvement and topics of interest that you would like to see in the future issues.

INTERACTIVE VISUALIZATION FOR CLINICAL SAFETY DATA REVIEW

Wei Wang, Eli Lilly Canada, Inc., Mary Nilsson, Eli Lilly, Rebeka Revis, Eli Lilly, Brenda Crowe, Eli Lilly

Introduction

As noted by several authors, e.g., Friedman, Furberg, and DeMets (2012), U.S. Food and Drug Administration (2010), Singh and Loke (2012), Wildfire et al. (2018) safety data tend to be more challenging to explore and interpret than efficacy data. Review of clinical safety data can be tedious and time consuming, akin to looking for a needle in a haystack. Safety evaluation often relies on the review of tabular displays of summary-level data in combination with a large amount of individual patient-level information in listing format. A typical review of safety data involves multiple iterations of evaluating summary-level information, requesting ad hoc analyses for more details or patient level information, and assessing this additional information. Fulfilling the requests for ad hoc analyses takes a lot of time and interrupts the review flow. Alternatively, teams can try to anticipate every possible analysis/display and create them right from the start. However, what is needed may be difficult to predict, and creating a large volume of analyses is unsustainable and takes time away from competing priorities.

A well-designed interactive visualization package provides a dynamic, review environment. Such a package, together with a multi-disciplinary review process, fits well with the nature of safety evaluation. With thoughtful display choices, the ability to drill down, conduct on-demand analyses, and instantaneously obtain additional detailed information, interactive visualization provides more effective presentation of complex data, increasing the likelihood of detecting key safety signals and improving the ability to make clinical decisions (Duke et al. 2015, Amit, Heiberger, and Lane 2008, Wildfire et al. 2018).

In this paper, we share our experiences and insights from developing and using interactive visualization of clinical safety data.

Using Interactive Visualization beyond Data Exploration

Many vendors and interactive visualization software users have indicated that they use interactive displays for informal/exploratory reviews only and that, when there are findings from the interactive review, they use traditional programing (e.g., SAS, R) to create a static display and do "proper" validation/QC on the computer program. While this practice of validating positive findings may be appropriate for exploratory efficacy analyses, we're not convinced that it is appropriate for review of safety data. For assessment of safety data, not finding something is just as important as finding something. Therefore, for interactive review of safety data, the proper validation or quality control needs to be in place to ensure the analyses are solid and appropriate. Secondly, we believe, with proper validation, peer review or quality control (QC), interactive visualization can be used for decision-making. Repeating the same analyses in SAS or R is redundant for a well-constructed and validated interactive package. Using SAS or R to repeat what's been done in a properly validated interactive review package is only needed if there are special formatting needs for a static display, for example, for publication purposes. It takes a lot of effort to create standard operating procedures for the validation of interactive visuals; however, the effort is worth it. Incorporating a solid validation and QC process into the construction of visualization packages allows the package to be used to make high-quality and prompt decisions.

Part of using interactive visualization beyond exploratory use is having the ability to document and retrieve the information. While there are various ways this can be accomplished, the "bookmark" feature that some software offers is very useful and often overlooked. Bookmarks are the saved settings for a given analysis, including all the relevant filters, visual selections, etc.

This feature allows users to quickly return to a previously created view of the data. One possible use of bookmarks is to have bookmarks for all the pre-planned analyses. This allows users to easily access the pre-planned analyses without manual navigation, or manual filter set up. Bookmarks are also very useful for traceability purposes. Important findings and data reviewed can all be "documented" in bookmark format and saved. Ensuring that the pre-planned bookmarks are correct can be part of the QC process.

Choosing the Right Graphical Presentation and Analytics

It is important that clinical decisions are based on high quality research and effective displays. As with any analysis, proper analytical methods need to be used and the displays should be well constructed in an interactive visual package.

We highly recommend the involvement of statisticians in the development of an interactive visualization package to ensure proper analytical/statistical methods are used. We have observed some examples of inappropriate methods in some interactive packages:

- The wrong denominator was used for the proportion of patients experiencing an adverse event.
- Numbers of patients reporting certain adverse events were compared between treatment groups instead of percentages (and the true denominators were not the same).
- Crude pooling was used for an integrated analysis of multiple studies with no option to do a properly stratified analysis.

These kinds of mistakes can be avoided by including a statistician in the development of the interactive package.

It's also important to choose an appropriate display for the intended purpose (Duke et al. 2015). For example, group means over time are not usually the best choice for assessing whether changes in labs/vitals are persistent versus transient. A spaghetti plot may be a better choice (PHUSE Workshop 2019 [11]). While a spaghetti plot can be created in a static format, it is typically better as an interactive display. As a static display it can be extremely crowded and hard to follow. Whereas, as an interactive display, zoom sliders/scroll bars can be included to better inspect a cluster of lines that might crowd together, and the hover-over feature can be used to see a line for a given patient.

As another example, a volcano plot (Figure 1) is a logical choice for general adverse event (AE) data, as it provides an assessment of the evidence of an imbalance between treatment and placebo and the magnitude of effect (PHUSE Workshop 2019 [11]). The volcano plot provides a simple view of "ranking" of the events from top right where the events have stronger evidence of an imbalance between the treatment vs placebo (i.e., low p-values) and have a higher magnitude of effect (e.g., large odds ratios) compared to the lower left. This is a suitable display that reviewers can use to identify events for further exploration based on the statistical evidence. While the volcano plot can be created in static format (Clinical Trials Safety Graphics Working Group 2013 [2]), it is vastly improved as an interactive display (Figure 1). The interactive version of a volcano plot no longer needs to rely on difficult-to-fit labels and a large number of colors and shapes to convey information. Users can hover-over a symbol to see the AE terms and additional information.

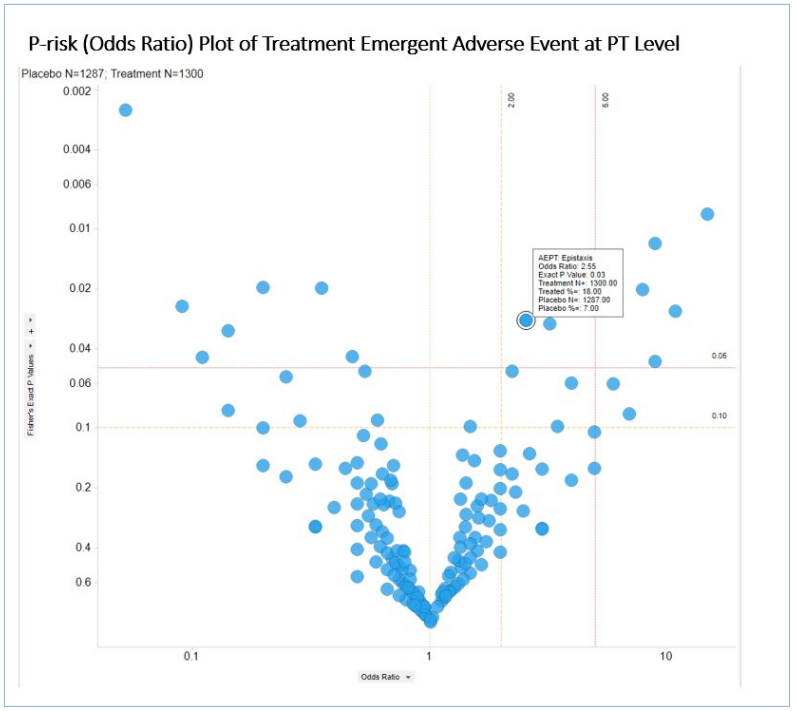


Figure 1. Snapshot of An Interactive Display for P-risk (Odds Ratio) Plot of Treatment Emergent Adverse Events at PT Level. Hover-over Shows Details about One Symbol

Statistical consideration can also impact which interactive features are the most desirable for certain interactive displays. For example, in the volcano plot above, statisticians may know that there is often a desire to review odds ratios and risk differences. An option can be incorporated into the interactive display allowing the choice.

Once the appropriate displays are identified, developers should ensure additional good graphical principles are followed (e.g., appropriate use of color, Duke et al. (2015)). Graphical displays are powerful tools to effectively and efficiently convey complex information. Well-designed graphs can help the audience better understand the objectives and results of clinical research. Poorly designed graphs can hinder the proper interpretation of data and distract reviewers from the key messages. They may even misrepresent the data.

Putting End-User (Reviewer) Needs First

It is important to get clinicians involved when developing an interactive review package. Understanding safety information is a cross-disciplinary effort and it is important to leverage the scientific expertise and medical judgment of people from different disciplines (e.g., physicians, epidemiologists, statisticians). The goal of the review package is to answer clinical questions. We have observed that some interactive packages provide access to clinical data but don't focus on answering important clinical questions. This may be due to the fact that the design of an interactive review package didn't start with a list of (or understanding of) the clinical questions, or the developer of the package did not have sufficient knowledge of clinical data and analytical methodologies. It is important to have both clinicians and statisticians involved in the development of an interactive safety review package. Clinicians provide input on the safety review needs and statisticians can design good analyses and displays to help the clinicians dig through massive amounts of information and identify important issues.

The following are some examples of how a good reviewer interface can be constructed.

It is helpful to design a guided review flow. For example, the review flow of adverse events could start with the display shown in Figure 1. The next step would be displays that provide additional details medical reviewers need to use for the events they pick. For example, for selected events, there is often a desire to understand the event severity, onset and duration. The appendix of the PHUSE white paper on the analysis and display for adverse events (PHUSE Adverse Events White Paper Team 2017 [7]) has a useful plot for the AE onset and duration display. In this plot, each event episode is plotted as a segmented colored line. Each line segment shows the start and stop of the event and color shows the severity of the event. A graphical patient profile is useful display for patient-level review. It allows a medical reviewer to look at events a patient experienced, the lab measures, concomitant medication uses, etc. on a common time axis. Reviewers select a group of patients with the events of interest in Figure 1 and look at their patient profile one at a time or from the AE onset and duration display and identify a patient to review. With this review flow in mind, a guided, intuitive and easy-to-navigate reviewer interface can be designed. For this kind of guided and stepwise review flow, it is important that users do not need to have go through too many steps (no more than three) to get to the most detailed information.

Another example is to have a list of analyses that contain hyperlinks or bookmarks (as noted in the section on Exploratory vs. Formal Analyses) to different analyses. This list helps reviewers focus on the a priori defined analyses. The reviewers can easily access these views through these links and not miss important information.

We suggest creating instructional text in the reviewer interface to help users learn how to conduct certain analyses on a page.

While interactive packages can be made simple and straightforward, training and coaching of clinicians is generally still necessary when they first use a package. We highly recommend that the package developer spends time going over the package with the clinician to make sure the analysis results and displays are properly interpreted. Clinicians need to know how to get to the information they need before they can conduct the review. In our experience, just-in-time training works better than training at the beginning of a roll-out of a package, as users will be more motivated to learn, and the training will be fresh in their memory when they conduct their review. It also helps to have hands-on training rather than just watching the instructor navigate.

When conducting the actual review, it is advantageous to set up focused cross-functional review meetings where clinicians can sit together with statisticians to go over the interactive review displays together. During these meetings, statisticians can also help clinicians with any questions they have on the interpretation of the display or get to the information they need for their questions.

Choosing the Right Software

Choosing the right software to invest in is very important. There are many visual analytics software packages. Some of the important differences in the features will not be appreciated until in-depth knowledge is gained about the software. In addition, company needs may change over time. Changing software is costly and sometimes difficult to do. Here are a few things to consider:

- The purchase price of the software itself is only part of the cost consideration. Very often, software has both an up-front fee and ongoing license fees. An important question to ask is what is included with these fees? Some software appears to be low-priced but may end up being much more expensive when things such as needed additional functionality, consultation, technical support and system setup service are added. It is also important to consider the operational costs. For example, open-source software is free but may come with increased programming complexity and maintenance cost. In addition, there is often a need to set up a special server for any large-scale use. The set up and maintenance cost of the server should be part of the cost consideration.

- The visual analytics needs are a key consideration for choice of software. For clinical trial data review purposes, we recommend software that has strong analytical power, and a decent variety of visualization options. While most visual analytics software supports common scientific graphics, many of them don't include sophisticated statistical analyses. This may be acceptable if the software has options to extend the core functionality with programming languages such as R, Python, JavaScript, HTML. These options allow users to extend the analytical power, customize and automate analyses or even embed visualizations into other web applications. If you have the necessary programming skills and would like to take advantage of these extensions, it is important to consider what extension options a software has.

- Consider the customizability of the reviewer interface. Many commercial packages are designed for users to create great-looking graphics but do not allow for easy customization of the reviewer interface. For example, some software allows for instructional text/images to be added in the reviewer interface and some do not.

- Do a thorough checking on the validity and reliability of the software. For example, in one of the software packages we checked, when there were multiple measurements at a single time point, the software randomly picked one of the measurements to plot on the graph instead of displaying all of them. Reviewers were not alerted that there were multiple measurements for a single time point.

- The resource cost of development of an in-house package is also an important point to consider. For example, a company with limited or no programming resources may want to consider off-the-shelf visualization packages (either commercial or open source such as Safety Explorer Suite [13] which is now bundled with the Hepatic Safety Explorer tool [6]). These predefined templates are typically built to handle data that are in a common format, e.g., Study Data Tabulation Model (SDTM). The purchaser/users do not need to know "programming" to use these packages. A company with extensive in-house programming resources may consider developing in-house packages that offer more flexibility and customization.

- Most interactive visualization software requires special knowledge. There are benefits to choosing software for which internal expertise is readily available. It is important to consider the time needed to train internal staff. A more complex software that offers sophisticated options might need specialized skills and require more training of internal staff.

- The choice of software can also be impacted by its purpose. For example, for a shared software repository where the goal is to build tools for cross-industry use, open-source software is a more favorable option because it is more widely available.

Sharing Interactive Visualizations Externally

Using interactive displays provides many benefits when communicating between pharmaceutical companies and regulators, payers, and even medical practitioners. When all parties are looking at data using the same sets of interactive displays, communication is clear and efficient. We have experience using interactive packages for advisory board meetings and received positive feedback from attendees. They communicated that the interactive displays helped them better understand our research results and that they considered it a very transparent way of sharing information.

Sharing externally is still challenging. Unlike static displays, interactive displays cannot be opened in common software packages such as Microsoft Office Word. Currently, many commercial software packages do not allow for a common sharing mechanism, which then requires both parties to have the same software. Some software allows for web-based sharing where users can go to a web site to access shared visuals. Since the on-demand recalculation and analyses is supported on the software server, there are still many concerns (e.g., data security) from both sponsor site and external parties to share or access information on the Sponsor's server.

A project team within an FDA/PHUSE collaboration (Rosario et al. 2012, PHUSE Working Group [10])is working on these challenges. The project team is "Best Practices for Interactive Analyses for Decision Making & Submissions" and resides within the Emerging Trends and Technologies Working Group. As part of this project, multiple options are being explored for sharing interactive displays/tools externally, such as in submissions (PHUSE Oct. Webinar Slides 2019 [9]). These options include:

- Giving external parties (e.g., FDA) access to a webserver

- Having an external party purchase the corresponding software

- Creating a third-party server

- Submitting a containerized app (Forrest Stroud) with the external party hosting on a virtual machine

- Including a stand-alone HTML interactive display as part of submission package

Cost is a factor for the first four options, as software needs to be purchased and/or a software engine/server needs to be created and maintained behind the scenes. Ensuring long-term integrity and access is a potential issue for the first three options. Processes and infrastructure would need to be created to ensure long-term integrity and access can be maintained. Firewall issues are a factor with the first two options. There are concerns (e.g., data security) from both sponsors and external parties to share or access information on each other's servers. Learning curves and training could be an issue for all five options, but less of an issue for displays/packages with simple interactive functionality. The last option, with a single display with simple interactive features, has been piloted and can be implemented today. The option, which involves the creation of an interactive display embedded in the submission document, allows the regulator to access the display without the need to install specialized software or accessing a server (PHUSE Oct. Webinar Recording 2019 [8]). The Hepatic Safety Explorer [6] was created in a manner that allows the tool to be exported to a fully functional version in a transportable html file that can be opened with any browser.

Sharing Knowledge and Engaging in Cross Industry Initiatives

It is always a good idea to learn from others people's experiences. In addition to searching for publications, books and internet resources on this topic, we encourage statisticians to be involved in cross-industry initiatives to learn available options and share learning. Reference to some of these cross-industry initiatives have already been made in previous sections.

Interactive visualization is an area that has recently developed very quickly. We are aware of several companies using interactive visualization in their clinical research. Regulators are also looking into this area. There are cross-industry working group that focus on using interactive visualization in clinician research. The PHUSE Working Group [10] has a visualization working group that has various sub-teams working on various aspects of interactive visualization for use in clinical research. PHUSE also has a publicly available code sharing repository that has example interactive safety review packages. The DIA-ASA Biopharmaceutical Safety Evaluation Working Group is developing a series of novel interactive safety graphic tools to enhance the ability of safety and clinical development professionals to identify and evaluate safety signals. Each will be made available as an open-source, non-proprietary application widely available to anyone interested in drug safety evaluation. Their Hepatic Safety Explorer tool [6] is already publicly available.

Interactive visualization is still relatively new and has great potential to improve how we review and communicate clinical research results. Regulators and industry need to work together to make interactive visualizations become a mainstream practice on how we review and communicate safety data.

Concluding Remarks

Interactive visualization can greatly improve safety data review of clinical trial data. We hope our experience can benefit those who are new to interactive visualization.

We believe that interactive displays are the future of how we share and communicate information. While many consider interactive displays to be limited to exploratory purposes, we advocate the use of properly validated and QC'd interactive displays that can support formal decision making. Statisticians are important in the development of interactive review packages to ensure the proper choice of the graphics and that the right analytic methods are used. When designing an interactive review package, it is beneficial to create a reviewer interface that is simple and intuitive. We recommend having a guided review interface with simple instructions. Currently, there are still many challenges with sharing interactive displays externally. Both industry and regulators have recognized the need to solve these issues. We encourage statisticians to get involved with cross industry efforts to increase the use of interactive displays and improve the environment for sharing interactive displays.

Acknowledgments

We are grateful to Jim Buchanan (Covilance LLC) and Rebecca Taha (Eli Lilly and Company) for their careful reviews of this manuscript.

References

All links were accessed November 16, 2019.

1. Amit, Ohad, Richard Heiberger, and Peter Lane. 2008. "Graphical approaches to the analysis of safety data from clinical trials." Pharmaceutical Statistics no. 7:20-35.

2. Clinical Trials Safety Graphics Working Group. Clinical trials safety graphics home page 2013. Available from *https://www.ctspedia.org/do/view/CTSpedia/StatGraphHome*.

3. Duke, Susan P, Fabrice Bancken, Brenda Crowe, Mat Soukup, Taxiarchis Botsis, and Richard Forshee. 2015. "Seeing is believing: good graphic design principles for medical research." Statistics in medicine no. 34 (22):3040-3059.

4. Forrest Stroud. Containerization. Available from *https://www.webopedia.com/TERM/C/containerization.html*.

5. Friedman, Lawrence M., Curt D. Furberg, and David L. DeMets. 2012. *Fundamentals of Clinical Trials*. New York: Springer.

6. Hepatic Safety Explorer. Available from *https://safetygraphics.github.io*

7. PHUSE Adverse Events White Paper Team. 2017. Analysis and displays associated with adverse events: Focus on adverse events in phase 2-4 clinical trials and integrated summary documents. *http://www.phuse.eu/documents//working-groups/cs-whitepaper-adverseevents-v10-4442.pdf*

8. PHUSE Oct. Webinar Recording. Sharing an Interactive Display in an FDA Submission Pilot 2019. Available from *https://www.youtube.com/watch?v=u26-A2NTj2k*.

9. PHUSE Oct. Webinar Slides. Sharing an Interactive Display in an FDA Submission Pilot 2019.

10. PHUSE Working Group. Available from *https://www.PHUSE.eu/PHUSE-working-groups.*

11. PHUSE Workshop. Safety Analytics Workshop 2019. Available from *https://www.phuse.eu/documents/working-groups/deliverables/safetyanalyticsworkshop-PHUSEcss2019-21959.pptx*

12. Rosario, Lilliam A, Timothy J Kropp, Stephen E Wilson, and Charles K Cooper. 2012. "Join FDA/PhUSE working groups to help harness the power of computational science." *Drug Information Journal* no. 46 (5):523-524.

13. Safety Explorer Suite. Available from *https://rhoinc.github.io/safety-explorer-suite/docs*/.

14. Singh, Sonal, and Yoon K Loke. 2012. "Drug safety assessment in clinical trials: methodological challenges and opportunities." *Trials* no. 13 (1):138.

15. U.S. Food and Drug Administration. 2010. Good review practice: Clinical review template. Manual of policies and procedures, *https://www.fda.gov/media/72472/download.*

16. Wildfire, Jeremy, Ryan Bailey, Rebecca Z Krouse, Spencer Childress, Britt Sikora, Nathan Bryant, Shane Rosanbalm, Emily Wilson, and Jack G Modell. 2018. "The safety explorer suite: interactive safety monitoring for clinical trials." *Therapeutic innovation & regulatory science* no. 52 (6):696-700. ■

NONPARAMETRIC AND UNSUPERVISED: NU-LEARNING FROM BIG DATA

Bob Obenchain, Risk-Benefit Statistics LLC, http://localcontrolstatistics.org

Background

Nonparametric methods are commonly described as "distribution-free." Unsupervised learning tends to be "model-agnostic." *NU Learning* approaches, by being clearly antithetical to traditional parametric-supervised model-fitting, offer unique opportunities to provide data-based insights that are both unbiased and truly objective.

NU-Learning Prototype

I have advocated a form of NU-learning, called "Local Control," for more than fifteen years [1-7]. The initial objective of "LC" Strategy is to obtain a *distribution* of unbiased "Local" Average Treatment Effect estimates (or "Local" Outcome-Exposure Associations) from *meaningful subgroups* [1] of experimental units (patients, etc.) Subgroups are formed by clustering units on their most relevant X-confounder characteristics. In this short article, I hope to stimulate interest in development of *NU* algorithms that are more efficient for truly Big Data. Specifically, I wish to pass a small "torch" forward to software developers who will, I hope, ultimately provide tools capable of cross-sectional analyses of many more than 100,000 patients. Readers interested in exploring a "subset" of their Big Data (say, at most 50K patients) can use my current R-package [5] to display key data-analytic *visualizations.*

Unsupervised Patient Matching vs Supervised Propensity Estimation

A string of key-concept papers [8-15] provides a strong foundation for both "approximate" patient *matching* and use of the *observed propensities* (treatment-choice fractions) within the resulting patient *clusters* to reduce bias in *comparative-effectiveness research.* Again, it's *not* necessary to risk fitting any possibly "wrong" global model (e.g. a logistic regression with interaction terms) simply to provide mere propensity "estimates." After all, algorithms for *clustering* and *matching* are the most widely used *NU* methods; better and better "K-Means" (or K-Medians) algorithms should emerge over time.

The current R-implementation of LocalControl Strategy [5] uses *hierarchical* clustering, even though such methods cannot "scale up" well to truly Big datasets. But hierarchical methods are usually good for "almost" big datasets where the number of patients (experimental units) is no more than roughly N = 50,000. Once a clustering tree (dendrogram) has been computed, it becomes efficient to monitor "bias-variance trade-offs" in "local" treatment effect-size estimation as the number of Clusters requested, K, is increased from 1 to 50, to 100, ...to at most, say, N/12.

Trade-Offs in Choice of Number of Clusters

When the smallest cluster formed contains at least 11 patients, many within-cluster statistics are (or should be) widely considered "publishable." For example, current CMS standards [16] for preserving patient privacy allow within-subgroup treatment effect-size estimates from "more than 10 patients" to be published in journals. If such detailed cluster-level information were made widely available for electronic download by publishers of research, all health-care researchers could evaluate this "supplemental information" and help in development of consensus views …especially if maximum cluster sizes are also restricted to, say, at most 20 or 30 patients. Authors and sponsors of sound research should be recognized for their data sharing efforts. In return, they retain their rights of access to and responsibility for preserving privacy of all patient-level information.

On the other hand, as illustrated below with Figures from a recent case study [4], using many *fewer* and much *larger* clusters (than 11 experimental units)

often "optimizes" clear Variance-Bias trade-offs in "local" effect-size estimation. But the extreme choice of using only K=1 "cluster" containing all N experimental units is *never* optimal; it provides only "one-size-fits-all" answers with minimum apparent variance but maximum true bias from confounded real-word data [8].

Visualizing Local Effect-Size Distributions

Estimates of "K" local treatment effect-sizes, *weighted* proportional to cluster size, can be displayed in a simple histogram, like Figure 1.

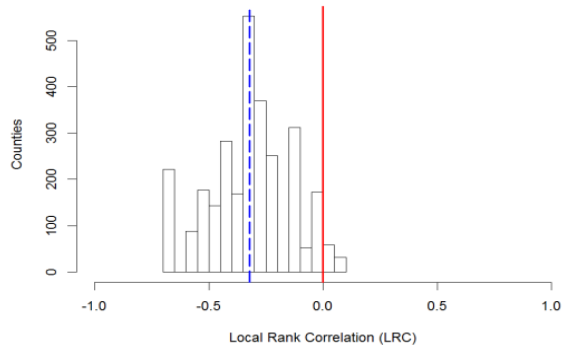


Figure 1. This is a histogram of local Spearman rank correlations between lung cancer mortality rates (y-outcome) and indoor radon exposure level within K = 50 clusters of 2,881 US counties [4]. The 3 X-confounders used to form clusters are percentages of county residents who (i) are over 65, (ii) currently smoke, and/or (iii) are obese. Since local associations are mostly negative here, note that local cancer mortality rates generally tend to decrease as radon exposure levels increase.

Next, displays like Figure 2 make it easy to literally "see" how initially stable or ultimately unstable the local "distribution" can become as K increases. Note that "Variance-Bias trade-offs" are illustrated using a sequence of box-and-whisker diagrams, each depicting a full "local effect-size distribution" as K is systematically increased. As outlined in the caption of Figure 2, an analyst literally "sees" that using only a few clusters (each much larger than 11 experimental units) optimizes this trade-off.

Permutation Test: Are X-confounders Ignorable?

A local treatment effect-size distribution is truly "meaningful" [1] only if its distribution is clearly different from the purely random distribution generated by random assignment of N units to K subgroups of same sizes (N_1, N_2, …, N_K) as the K observed clusters. A two-sample Kolmogorov-Smirnov "D-statistic" can then be used in a permutation test, but its "p-value" must be simulated because there are *many* within-cluster ties; see Figure 3 (next page). In any case, it is rather easy to determine whether this p-value is or isn't less than 0.01. My recent presentation at MBSW [6] gives five case-study examples.

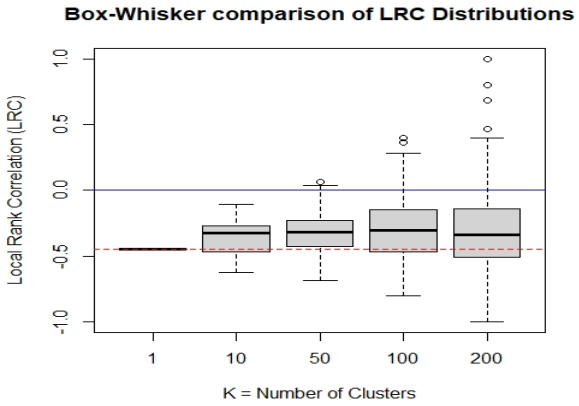


Figure 2. The most obvious effect of increasing K is that successive "local" effect-size distributions become more and more "spread out" vertically. Corresponding changes in (negative) LRC medians are not monotone; they start increasing towards zero for K = 10 & 50, reach their maximum at K = 100, and then start decreasing at K = 200. Ultimately, Median values tend to bob Up-and-Down for K > 200 (not shown) while the variability in LRC estimates increases monotonically. All of this suggests that roughly K = 50 is "optimal."

Systematic Sensitivity Analyses

How sensitive is the location and shape of the distribution of Local Effect-Size estimates to different clustering parameter-settings? To answer this question, analysts using LC Strategy [5] must first decide which of the available X-confounder characteristics are included, and which are excluded, in the process of clustering experimental units. All potential Y-outcome measures (the "left-hand side" variables in model-fitting approaches) as well as the primary treatment indicator or exposure-level measure must be excluded from consideration in the LC process of forming clusters.

The analyst also controls selection of clustering algorithm. The default algorithm [5] is *ward.D* but *diana* or *complete* linkage are also available. However, *single* linkage is *not* recommended!

Prediction of Local Effect-Sizes

LC strategy *deliberately* separates *estimation* of "local" effect-sizes via (nonparametric) *unsupervised* learning from their *supervised prediction*. Once random-permutation testing *confirms* that the X-confounders used in clustering are *Highly Unlikely* to be ignorable, *prediction* of local effect-sizes using *supervised* methods can be attempted with reduced concern about being misled by over-fitting.

Furthermore, when only a few of the most relevant X-confounders are used in clustering, the *within-cluster* variability in these X-confounders is essentially minimized. In this way, key *between-cluster* X-variation is preserved for use in *predicting* the corresponding variation in both y-outcomes and in treatment / exposure relationships. My experience is that *better predictions* are then much more likely to result!

In this optional final phase of LC strategy, the "stretch goal" is to reveal the extent to which local effect-size estimates are heterogeneous across clusters. In other words, the objective is to show that LRC estimates are predictable *fixed* effects rather than homogeneous (unpredictable) *random* effects.

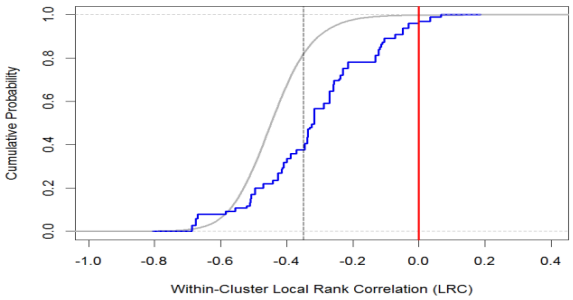


Figure 3. Two empirical Cumulative Distributions Functions for LRC estimates are shown here. The eCDF with 50 steps depicts the observed LRC distribution, while the "purely random" eCDF formed using 1,000 independent replications (each using 50 clusters of the same sizes as the observed clusters) looks quite smooth. The observed K-S D-statistic [4] of +0.454 at roughly LRC = –0.35 (vertical line) seems "gigantic." However, it's p-value cannot be "looked-up" in some standard table because both distributions are *discrete* (contain thousands of tied values.) A valid p-value can only be simulated using *another* 1,000 independent replications. Here, the simulated p-value is less than 0.001 because the largest of 1,000 simulated NULL "D"-values is < 0.22. This clearly suggests that the 3 X-confounders used to form the LRC distribution for 50 clusters (Figure 1) are Not Ignorable and, in fact, should be meaningful predictors of LRC variation across US Counties.

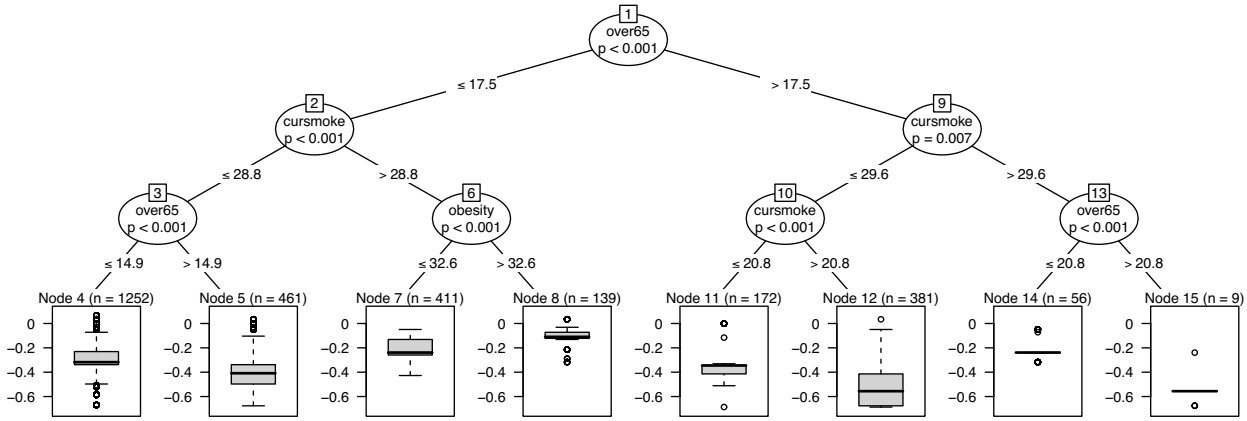


Figure 4. In this small *party tree predictive* model [17,18], final node 4 is quite large (1,252 of 2,881 US counties), and its local distribution of LRC estimates is much like the full distribution for all 2,881 counties. All 7 node splits shown have p-values not only < 0.001 but also less than 0.00015. This ultra-simple model (R^2= 0.472) explains slightly less than half of the total across-cluster variation in LRC estimates. These seven splits thus represent Heterogeneous Effects that provide new insights regarding potentially causal relationships detailed in [4].

Because clusters commonly vary considerably in size, it is essential to attach weights to individual local effect-size estimates when fitting traditional "parametric" across-cluster models. Our experience is that simply using weights directly proportional to cluster sizes is both realistic and robust. All available predictor variables, including radon exposure level itself, can then be used in attempts to predict the observed distribution of effect-sizes.

In Figure 4, we illustrate that use of Recursive Partitioning (nonparametric supervised learning) can avoid weighting issues by again relying upon within-cluster tied estimates.

Final Remark

I hope some readers now feel motivated to run the *demo(pci15k)* or *demo(radon)* examples of LC Strategy [5]. Readers could also modify the R-code within either of the above demos or in my LC Vignette [7] to gain new insights into their own "almost" large cross-sectional datasets.

References

1. Obenchain RL. "Identifying Meaningful Patient Subgroups via Clustering - Sensitivity Graphics." *Biopharmaceutical Section: Proceedings of JSM*, 2006.

2. Obenchain RL. "The local control approach using JMP®." Chapter 7 of: *Analysis of observational health care data using* SAS, ed. D. Faries, A. Leon, J. Maria-Haro, R. Obenchain, 151–192, Cary, NC, SAS Press, 2010.

3. Obenchain RL, Young SS. "Advancing statistical thinking in health care research." *Journal of Statistical Theory and Practice*, 2013; 7, 456-469.

4. Obenchain RL, Young SS, and Krstic G. "Low-level radon exposure and lung cancer mortality." *Regulatory Toxicology and Pharmacology,* 2019. doi: https://doi.org/10.1016/j.yrtph.2019.104418

5. Obenchain RL. LocalControlStrategy: R-package for robust analysis of cross-sectional data. version 1.3.2 https://CRAN.R-project.org/package=LocalControlStrategy, 2019.

6. Obenchain RL. "Are X-confounders Ignorable? ...a Permutation Test using Random Patient Clusters" MBSW. 2019. link to presentation.

7. Obenchain RL. LocalControlStrategy: R-package Vignette. 2019. http://localcontrolstatistics.org/other/LCstrategy_in_R.pdf

8. Cochran WG. "The effectiveness of adjustment by subclassification in removing bias in observational studies." *Biometrics* 1968; 24, 205-213.

9. Rubin DB. "Bias reduction using Mahalanobis metric matching." *Biometrics* 1980; 36, 293-298.

10. Rosenbaum PR, Rubin RB. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 1983; 70, 41-55.

11. McClellan M, McNeil BJ, Newhouse JP. "Does More Intensive Treatment of Myocardial Infarction in the Elderly Reduce Mortality? Analysis Using Instrumental Variables." *JAMA* 1994; 272, 859-866.

12. Lunceford JK, Davidian M. "Stratication and weighting via the propensity score in estimation of causal treatment effects: A comparative study." *Statistics in Medicine* 2004; 23: 2937-2960.

13. Bang H, Robins JM. "Doubly robust estimation in missing data and causal inference models." *Biometrics* 2005; 61: 962-973.

14. Ho DE, Imai K, King G, Stuart EA. "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference." *Political Analysis* 2007; 15, 199-236.

15. Stuart EA. "Matching Methods for Causal Inference: A Review and a Look Forward." *Statistical Science* 2010; 25, 1–21.

16. Standard Form CMS-R-0235L. "Instructions for Completing the Limited Dataset Data Use Agreement, Section 8.a." http://www.cms.gov/ Medi care/CMS-Forms/CMS-Forms, 2012.

17. Hothorn T, Hornik K, Zeileis A. Unbiased recursive partitioning: a conditional inference framework. *J. Comput. Graph. Stat.* 2006; 15 (3), 651–674.

18. Hothorn T, Hornik K, Zeileis A. *party*: A Laboratory for Recursive Partytioning. R-Vignette, 2006. ■

QUANTUM COMPUTING AND STATISTICS

Sergei Leonov, CSL Behring

Quantum computing has received a considerable interest in recent years in a number of diverse areas, from financial engineering to cryptography to artificial intelligence. New quantum computing algorithms, including quantum Monte Carlo, are developed in parallel with building first functioning quantum computers. While the theory of quantum information and various theoretical aspects of quantum computing have attracted statisticians and probabilists for decades, the examples of practical applications are still rather rare: despite a few well-known quantum algorithms, such as Shor's integer factorization algorithm and Grover's search algorithm, the number of specific problems for which quantum computing has proved its advantage is limited.

The theoretical power advantage that quantum computing holds over conventional computing primarily relates to two principles of quantum mechanics – superposition and entanglement. For a nice non-technical introduction to these concepts, see Gamble (2019). A website of D-Wave Systems, the manufacturer of the first commercially available quantum computer, covers a wealth of information on quantum computing and includes tutorials which are designed to be accessible to both technical and non-technical audiences; see *https://dwavesys.com*. For technical details, see monographs by Nielsen and Chuang (2010) and Wittek (2014) and an overview article by Wang et al. (2016). Quantum annealers which are developed by D-Wave Systems and universal (general purpose) quantum computers, with Google and IBM leading the charge, are not competitors. While relying on the same concepts of quantum mechanics, they are useful for different sorts of problems and utilize different types of architecture.

Superposition: For conventional computing, the basic unit, a bit, exists in one state at a time, and this state is deterministic, either 0 or 1. Rather than storing information using bits represented by 0s or 1s, quantum computers use quantum bits, or *qubits*, to encode information as 0, 1, or both at the same time. Superposition means that the qubit exists in two states at one time, and these states are probabilistic, adding up to 1. Therefore whereas classical computing is limited to manipulating binary bits in a deterministic stream, quantum computing can manipulate vast data sets simultaneously in a probabilistic space. The magnitude of this difference is illustrated by the number of states a quantum computer can represent. The latest D-Wave computer has 2,000 qubits, which could exist in a superposition of as many as $2^{2,000}$ or about 10^{600} quantum states. For comparison, it is estimated that there are about 10^{80} atoms in the known, observable universe.

Entanglement: The second advantage of quantum computing is entanglement, which means that individual qubits can interact directly with each other even at great distances, altering each other's states simultaneously without intermediate causal connections. If the reader finds the concept of entanglement rather counterintuitive, be assured that you are not alone: Albert Einstein once described entanglement as "spooky action at a distance".

As an example of an optimization problem well suited for quantum computers, take a minimization problem with many local minima. Traditional global optimization methods such as simulated annealing utilize ideas from thermodynamics to implement hill-climbing and avoid being stuck in a local minimum; see Bohachevsky et al. (1986). On the other hand, quantum annealing algorithms can use computational shortcuts such as quantum tunneling which are not available in conventional computers. To visualize the difference between simulated and quantum annealing, think of a mountainous terrain with many hills and valleys. While searching for the deepest valley, simulating annealing algorithms will move from a low valley to a lower one over the surface of the Earth, one small step at a time. On the contrary, quantum annealing makes it possible to move simultaneously from several not-so-deep valleys in the direction of the deepest one via tunneling, i.e., through the hills.

Harnessing the potential power of quantum computing will require significant advances in technology, with serious hardware challenges to overcome. One particular challenge is providing a super-cool environment: since even minor interactions with the external world

may change the state of a quantum system, a stable quantum system can only exist at temperatures close to absolute zero (0°K = -273.15°C = -459.67°F). For instance, the latest generation D-Wave system operates at 15 millikelvin; see *https://www.dwavesys.com/tutorials/background-reading-series/introduction-d-wave-quantum-hardware*. Given the physical limits involved, developing software and interfaces capable of reliably working with quantum models is difficult. While it will take time and effort to apply quantum computing to solving various practical problems, even partial success would represent a "quantum leap" in computing power; so the effort is well worth it.

On October 23, 2019 Google announced reaching "quantum supremacy" - solving a problem which cannot be solved by a standard computer within a reasonable amount of time. A dramatic increase in speed was demonstrated: a specific problem related to random number generation was solved on a Google Sycamore processor with 53 qubits in 200 sec while the equivalent task for a state-of-the-art classical supercomputer would have taken approximately 10,000 years; see Arute et al. (2019).

ASA Scientific Interest Group on quantum computing

In 2016, members of ICON Innovation Center, which at that time included the author of this article, got involved in discussions with a group of researchers from Lockheed Martin Corporation, exploring statistical problems which can be solved on a quantum computer. Lockheed Martin became the first customer of D-Wave Systems in 2011 with the purchase of a 128 qubit D-Wave quantum computer; see *https://www.dwavesys.com/press-releases/d-wave-systems-announces-multi-year-agreement-lockheed-martin.*

In the Spring of 2017, ICON was granted access to D-Wave 1000-qubit quantum computer of Lockheed Martin through its affiliates program. D-Wave computer is a quantum annealer which is designed to solve quadratic unconstrained binary optimization problems (QUBO):

$$\boldsymbol{x}^* = \text{Argmin}_{\boldsymbol{x}} \, [\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{h}^{\mathrm{T}}\boldsymbol{x}],$$

where elements of an $n\times 1$ vector $\boldsymbol{x}$ are either -1 or 1, $\boldsymbol{Q}$ is an $n\times n$ symmetric matrix, and $\boldsymbol{h}$ is an $n\times 1$ vector. A popular Ising model of ferromagnetism provides a popular example. Other examples of practical applications which can be reduced to a QUBO problem include a travelling salesman problem, graph coloring, binary integer linear programming, facial recognition and space mission planning.

Quantum computing could prove useful when designing clinical trials. For example, statistical methods of trial design that fuse combinatorial and model-based optimal experimental design techniques are developed that may substantially reduce the number of sites and patients needed to select the best treatment combinations for targeting multiple cancer types and multiple biomarkers. The fusion of the two approaches is promising but can be computationally challenging, which motivated us to explore quantum annealing algorithms while running iterative numerical methods for optimal design construction; see Fedorov and Leonov (2018).

In March 2017, Lockheed Martin, ICON Innovation Center and George Washington University (GWU) Department of Statistics co-organized a workshop on quantum computing and its application in drug development which took place at GWU in Washington, DC. The workshop was attended by more than 50 participants from academia, government, and industry; see *https://statistics.columbian.gwu.edu/workshop-quantum-computing-and-its-application*. It featured overviews of quantum computing, talks on quantum algorithms and their links with statistics, as well as case studies and a round table discussion.

Following the 2017 GWU Workshop, a number of sessions on quantum computing were organized at statistical conferences, which included JSM 2017 (Baltimore), CEN-ISBS 2017 Joint Conference on Biometrics and Biopharmaceutical Statistics (Vienna, Austria), JSM 2018 (Vancouver), and JSM 2019 (Denver). Speakers at these sessions represented academia (GWU, University of Toronto, Cardiff University, University of Calgary), government (National Institute of Standard and Technology (NIST), Oak Ridge National Laboratory) and industry (Lockheed Martin, ICON, Nokia Bell Labs, QxBranch, ProteinQure). Such a broad representation exemplifies the cross-industry interest that will hopefully move this potentially revolutionary computing technology from theory to practical applications.

In 2018, a Scientific Interest Group (SIG) on Quantum Computing in Statistics and Machine Learning was organized within the ASA Section on Statistical Computing. Valerii Fedorov led the organizational efforts and became the first Chair of the SIG. Peter Wittek (University of Toronto) and Sergei Leonov were inaugural Program Chair and Secretary, respectively. Peter and Sergei accepted the roles of Chair and Program

Chair in 2019. In September 2019, the SIG organized a webinar on quantum computing which featured a presentation by Mark Fingerhuth (ProteinQure) and was attended by more than 50 participants. Among the most recent plans of the SIG are the organization of a session at JSM 2020 (Philadelphia, August 2020) and a workshop on quantum machine learning at the Banff International Research Station (Canada, July 2020). For more details about the SIG, see *https://community.amstat.org/quantum-sig/home.*

It is with great sadness that we share the news that our friend and colleague Dr. Peter Wittek went missing on September 29, 2019 while climbing in the Himalayas. Peter, a world-renowned expert in quantum computing and the author of the seminal book on quantum machine learning, served as the Chair of the SIG in 2019. A skilled mountaineer, Peter has climbed the world's most famous peaks including Mount Kilimanjaro in Africa and Aconcagua in South America. In late September he was swept away in an avalanche while climbing Mount Trishul in the Indian Himalayas.

This short article is dedicated to Peter's memory.

References

1. Arute, F. et al. (2019). Quantum supremacy using a programmable superconducting processor. *Nature,* 574, 505–511.

2. Bohachevsky, I., Johnson, M.E., Stein, M.L. (1986). Generalized simulated annealing for function optimization. *Technometrics*, 28(3), 209–217.

3. Fedorov, V.V., Leonov, S.L. (2018). Combinatorial and model-based methods in structuring and optimizing cluster trials. In: Beckman R, Antonijevic Z. (Eds.) *Platform Trials in Drug Development: Umbrella Trials and Basket Trials. Chapman & Hall/CRC*, pp. 265-286.

4. Gamble, S. (2019). Quantum computing: what it is, why we want it, and how we're trying to get it. In: *Frontiers of Engineering. Reports on Leading-Edge Engineering from the 2018 Symposium.* The National Academies Press, Washington, DC, pp. 5-8. *https://www.ncbi.nlm.nih.gov/books/NBK538701/*

5. Nielsen, M. A. , Chuang, I.L. (2010). *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, Cambridge.

6. Wang, Y., Wu, S., and Zou, J. (2016), Quantum annealing with Markov chain Monte Carlo simulations and D-Wave quantum computers. Statist. Sci., 31(3), 362–398.

7. Wittek, P. (2014). *Quantum Machine Learning. What Quantum Computing Means to Data Mining*. Elsevier, Amsterdam. ■

CRAIG'S LIST: QUOTES AND SAYINGS FOR LIFE AND LEARNING

Craig Mallinckrodt, Biogen

Introduction

In 2017 my long-time employer offered a voluntary early retirement plan. Although I was happy and rewarded there, I was ready for change. I accepted the early retirement offer. In the months leading up to my departure, I reflected on my career and gathered input from colleagues regarding what I had done that mattered most to them.

Craig's List is the product of that reflection - a collection of sayings. Few are my original thoughts, and I can only rarely cite the source for the others. Moreover, simple one-liners cannot fully explain what to do and how to do it. Nevertheless, I share these sayings hoping that they will guide you as usefully as they guide me.

The sayings are arranged into themes. Some are self-explanatory. For others, additional comment is provided.

Technical acumen

1: The difficulties lie not so much in knowing the principles, but rather in putting them into useful practice.

This saying describes a fundamental challenge statisticians in the pharmaceutical industry face. Understanding this challenge provides insight into how we can continue to improve as our careers progress.

2: If a result doesn't look right, act like it isn't.

3: Things only become obvious after they become obvious.

Sayings 2 and 3 instruct us to work hard to get things right. In retrospect, solutions may seem obvious, but before we find them they are not easy and we need to be prepared for a hard search.

4: The quest for perfection gets in the way of good enough.

5: I don't care what you do or how you do it, so long as it has the highest probability of giving me the right answer.

We shouldn't choose methods to advance the cause of those methods, or to mindlessly adhere to tradition. Our choice should be based on what works best.

6: Don't talk to me about averages - because if you put your head in the freezer and your butt in the oven, on average you should be comfortable.

This quote came from my father. I used it to explain to diverse audiences the importance of understanding variation in outcomes. The quote was useful because it put a complex idea into simple terms that everyone could understand – and most thought was pretty darn funny.

Philosophical approach to work and career

7: Keep your eye on the football.

The "football" is the shape made by the overlap of circles in a Venn Diagram. Figure out what is important to your team. Figure out what you like to do. Spend as much time as possible where these two overlap.

8: Be a multiplier.

If you work as effectively as you can, you contribute 1X. If you can also help 10 others to work 10% more effectively your contribution is 2X.

9: Take what you find and make it better.

10: Some look at the way things are and ask why. He looked at the way things ought to be and asked why not.

Saying 9 is from a founding family member of my long-time employer. Ten is from Senator Ted Kennedy's eulogy of his brother Robert. These sayings remind us that it is easy to complain, but it doesn't help.

11: You can't eat the apple in one bite.

12: The man who moves a mountain begins by carrying away small stones - Confucius.

Sayings 11 and 12 advise us to break down big tasks into component pieces so that the overall task is not too intimidating to begin.

13: The best time to plant a tree was 30 years ago. The next best time is today.

General advice

14: Don't compare salaries or promotions. These comparisons lead to greed and jealousy, and those won't help you get a raise or a promotion.

15: Success requires doing things right and doing the right things.

16: Greatness isn't one thing, it is many little things all done well.

17: Practice makes ~~perfect~~ permanent.

The old saying is that practice makes perfect, but it doesn't, unless the practice is perfect. Practice engrains actions, responses, habits, etc. whether they are perfect, good, or bad. Therefore, to improve we must practice diligently.

Perspective

18: Good judgement comes from experience, and experience comes from bad judgement.

19: Quitting is a permanent solution to a problem that is likely temporary.

20: Attitude and approach are more important than aptitude.

21: When one door closes, you need to open others.

22: I didn't get the things I wanted. I got the things I needed instead.

These sayings remind us that our careers have ups and downs. We make mistakes. We have failures. We get discouraged. How we respond in these down times says more about our character and our prospects for success and happiness than how we respond in good times.

Performance and productivity

23: Many people think that giving a good presentation is glamorous, but nobody thinks the preparations required to give a good presentation is glamourous. (Ditto for writing good papers.)

This is a variation on the quote that the will to win means little without the will to prepare.

24: When I started, I thought time management was the most important skill. Now I think it is attention management.

25: Make time to think, every day. Take a walk outside, every day.

26: I can do 12 months of work in 11 months, but I can't do 12 months of work in 12 months.

Today's work and social environment pose challenges to our ability to sustain long, successful, and happy careers. One of the most useful things I did was to study how to work efficiently and creatively in today's highly distractible work environment. I encourage you to do the same.

27: Find the 10%.

A 10% increase in the efficiency of your work may not seem like much, but it yields the equivalent of 25 extra days each year to do other things.

Learning

28: You have the perfect background for this position. The other 90% of what you need to know you can learn on the job.

To sustain productive and happy careers we must continue to learn on the job.

29: In learning you will teach, and in teaching you will learn.

30: No one learns as much about a subject as one who teaches it.

31: Tell me and I forget, teach me and I may remember, involve me and I will learn.

Quotes 29-31 point to active learning and teaching for individual and organizational learning. Specifically, identify an important problem, study it or conduct research to solve the problem, and then teach what you learned to others. It's a win-win for you and your organization.

32: A student was given a mentoring opportunity because having someone leaning on her would help her to stand even steadier.

I cannot say how much those I mentored gained from the experience; I know that I gained a lot.

Leadership

33: The mediocre leader tells. The good leader explains. The superior leader demonstrates. The great leader inspires.

34: Your most important task as a leader is to teach people how to think and ask the right questions so that the world doesn't go to hell if you take a day off.

35: Leaders should influence others in such a way that it builds people up, encourages and edifies them so they can duplicate this attitude in others.

Ethics and behavior

36: The night before your drug hits the market, it is comforting to know that you did your best to find the right answers.

37: When you are trying to figure out what to do, ask yourself, how would my family feel about my actions if they read about it on the front page of USA Today.

Ethics is a complex topic. However, if you have done your best, and those whose opinion you value are proud of your actions, you must be on the right track.

38: On your last day at work, your colleagues will not recall how many papers you authored or how many submissions you did. They will remember how you made them feel.

We can't go through our careers with the only goal of having people like us, but if people like us we are more likely to achieve our career goals – and be a lot happier.

Priorities

39: If work is the most important thing in your life, you should work on your life.

For many, work will be more rewarding if it is part of a balanced life.

40: The days are long but the years are short.

Use your time wisely, it will pass quickly. This applies to work, family, hobbies, etc.

41: Answer these three questions: Am I working hard; am I working smart; am I working well with others?

If all three answers are yes, you are on track for success. If one or more answers is no, you need to make changes. If you don't know the answer to one or more, you need to seek feedback.

42: Be sure to enjoy this day.

This advice came from my wife Donna on the morning of the most meaningful presentation of my career. The preparations were exhausting and I'd lost perspective. Donna's advice helped me to appreciate how lucky I was to be part of such an important moment. Fatigue and burden became appreciation and opportunity. Our emotions and our attitude are powerful forces. Learning how to focus them in positive directions leads to greater success and satisfaction.

> "THE NIGHT BEFORE YOUR DRUG HITS THE MARKET, IT IS COMFORTING TO KNOW THAT YOU DID YOUR BEST TO FIND THE RIGHT ANSWERS."

A final word

43: Success is a journey, not a destination.

This often-quoted line has several useful contexts. First, think of goals as mileposts on your journey, not the destination. The goals are the means to achieve a greater end. Second, the route is more meaningful than the destination. Two people can get to the same place, achieve the same goal, but the journey could be much tougher for one, depending on where they started and what obstacles they encountered along the way. Third, the value or reward in a journey is derived not from the destination, but from what is gained along the way.

Say that you set the goal of running your first marathon. You train diligently. On race day, you execute your pacing plan and you finish the marathon! You earn a coveted finishers medal and you celebrate your achievement. Ten years later, what from this experience will be important?

You will likely have stored your medal in some obscure location. You will hardly remember the post-race celebration. But, you will remember the motivation and dedication it took to achieve your goal. You will remember that to improve your ability you had to work hard and work smart. Most of all, you will remember the satisfaction and pride from having done your best. Awards and rewards are great, but they fade over time. Memories, experiences, friendships, satisfaction, being part of something meaningful, these endure, and these are likely useful mileposts for your journey. ■

A CAREER AS A PHARMACEUTICAL BIOSTATISTICIAN

Steven Snapinn, Independent Consultant

This past October the department of biostatistics at the University of North Carolina held a 70th anniversary celebration and, as an alumnus of the department, I was asked to give a presentation. Events like that encourage one to reminisce, and I admit that I did indulge regarding my time in Chapel Hill and my career since then.

I had just begun a career as an engineer when I decided to enroll in the PhD program in biostatistics at UNC. The move to Chapel Hill was somewhat of a culture shock, in particular the move from an apartment on West 115th Street in upper Manhattan to a room in a house in the middle of 35 acres of pristine forest about 10 miles out of town. At UNC, I immediately realized that this was a great choice for me – I had always wanted to apply my math skills to advance public health, and biostatistics was clearly the perfect way for me to do that. The faculty were great, the curriculum was great, and I'd like to particularly thank Jim Knoke, who wasn't one of the best-known faculty members, but who was a wonderful dissertation adviser.

Since graduating in 1983 I've spent my career in the pharmaceutical industry, primarily at Merck and Amgen. In my years at Merck I learned how to be a pharmaceutical statistician, and I had the privilege to work on several landmark clinical trials that changed cardiovascular medicine. I also learned that statisticians in the industry shouldn't stop doing statistical research, and that there's a never-ending stream of statistical problems that need new or better methods. After moving to Amgen, I took on more of leadership role and I tried to instill that message in my team. And I've tried to maintain academic relationships, with UNC and other programs, where, among other things, I participated on several dissertation committees. More recently, I got to experience what life is like in a small biotech company and enjoyed the camaraderie of that environment.

I've always been interested in a wide range of statistical issues, touching on many but not going particularly deeply into any of them. But I've recently come to realize that these issues tend to share a couple of common themes, survival analysis and dichotomization of a continuous variable, and in some cases the issues overlapped both themes. For example, right at the nexus is an issue that dates to my work on cardiovascular outcomes trials at Merck. It became clear to me that identification of nearly every clinical outcome (all-cause death being the notable exception) involves some degree of judgment. The standard approach is to define criteria or employ an adjudication committee to apply that judgment with some objectivity and classify the potential endpoint as true or false. However, this approach ignores the degree of certainty behind that adjudication decision, and innovative methods that incorporate the degree of certainty can be considerably more efficient. In fact, I've come to realize that the tendency to dichotomize continuous variables in order to define "responders" is perhaps the most insidious and damaging issue in the statistical analysis of clinical trials. Another issue related to the theme of survival analysis involves determining which measure of treatment efficacy is best for assessing clinical meaningfulness: a relative measure like a hazard ratio, or an absolute measure like a risk difference or a difference in medians. Most recently I've become interested in a fundamental question when the hazards are nonproportional: what, in fact, is the null hypothesis being tested?

It's the never-ending stream of issues like these, as well as the knowledge that one's efforts lead to the development of innovative new medicines, that make a career in the pharmaceutical industry so rewarding. ■

BUSTING SOME MYTHS ABOUT STATISTICAL LEADERSHIP

Gary Sullivan, Espirer Consulting

Have you ever offered a new analysis approach in a cross-functional meeting only to be ignored or voted down? Did you ever leave the room frustrated because the decision-maker wanted to use the same method or study design even though better, novel approaches existed? Have you watched as your collaborators – physicians, regulators, scientists, engineers – drive the strategic discussion and emerge as project leaders? Have you become resigned to the idea that the role of the statistician is one of support, to provide study designs, analyses, and reports when requested? If your response is, "Yes, but I don't know what to do!" The answer is to learn how to lead – now!

When I look back on my career I can think of those instances where I left opportunities on the table or had good ideas that I couldn't convince others to adopt. Early in my career, I proposed an idea to my senior director (a non-statistician) to use statistically designed experiments to establish operating ranges and show robustness in the transfer and scale-up of processes from development into manufacturing. (This was several years before the FDA's 21st century initiative that made this more of a requirement.) But I didn't have the skills to take the idea forward. I lacked skills in networking, influence, communication, and strategic thinking, to name a few. Only later would I realize investing in those leadership skills would help me drive change and deliver greater value to the organization. I had certain beliefs about my role as a statistician and eventually realized that thinking was a self-imposed limitation.

So, with apologies to the TV show, I'd like to do some "mythbusting" to reset some beliefs statisticians hold, open possibilities for greater impact, and allow for more challenging & rewarding career paths. I'll finish with some ideas on leadership study and upcoming training opportunities that can set you on the path toward stronger leadership. Onto the "mythbusting" …

Myth #1: Leadership development is only for the highest potential statistical talent**.**

The truth is that ALL statisticians who want to consistently grow, impact and advance will increase their rate in these areas by investing in leadership development. That means statistical managers, technical supervisors and individual contributors.

Let's start with individual contributing statisticians. You might say, "Why leadership? I'm technical. I apply methods. I design/analyze/interpret data. I innovate. Leadership? Not my job." Back to the opening questions … let's suppose you believe a newer method or approach should be applied to design or analyze a study. If it's a small study, maybe you get your way because you are the expert. But if the study impacts a larger project that involves significant resources ($$, people, patients, materials), timeline implications, regulatory considerations, and operational changes, you've got a

whole host of others you will need to convince including physicians, senior scientists/engineers, operations directors, project managers, marketing directors, and several other functional leaders & stakeholders. This will require some of the following: business/scientific/regulatory knowledge, networking, effective communication, operational understanding, coalition building, negotiating, political savvy, change management, effective communication. In other words, leadership skills.

How about administrative managers whose responsibilities are supervising large (15 or more) groups of people? Included in this group are technical supervisors - those with both people AND technical responsibilities (perhaps the most difficult role). Skills required for "people" managers include the ability to strategize, communicate, build trust, delegate, challenge, motivate, reward (and retain!), influence, resolve conflict, and prioritize. How do managers learn to do all these things? Some copy how it was done for them. Some rely on instinct. Some take a course in supervisor training. Some rely on mentoring. Some think about what it

takes to be a true leader of people and work to achieve it. Where do you fall? Most importantly, are you working to get better?

Myth #2: The only way to improve your leadership skills is thru experience.

The truth is that experience is a great teacher of leadership, but there are additional ways to accelerate your leadership growth and development.

Even the most effective statisticians don't get to a point of strong leadership until the second half of their career. Why? Experiential learning takes time. Wouldn't it be helpful if you could accelerate that learning and achieve that level of peak effectiveness five or ten years sooner? Those experienced, senior statisticians who have achieved this level can point out several critical points in their career where certain experiences - both good and bad - helped them learn critical leadership skills and ultimately made them better leaders and more impactful statisticians. I would guess that every one of them would have liked to learn and know those skills earlier in their career. Can that be done? Yes. In the words of Vince Lombardi, a famous American football coach, "Leaders are made, not born." Recently, I had a new pharmaceutical statistician share with me their frustration in not being able to influence as effectively as their senior counterparts when they felt more technically expert in certain areas. The good news for this statistician and others is that these skills can be learned, and not just through experience.

Myth #3: Statisticians are trained to support, not lead.

There is some truth to this, but it does NOT mean you cannot lead. The truth is that, through your technical training, you already have some important foundational leadership skills which position you to be effective leaders. And, more importantly, the industry needs statistical leaders.

As statisticians, you already have some leadership skills in your toolbox including problem solving, learning agility, and the ability to process complex information. In addition, you are collaborative, objective, studious, think deeply, and tend to have high integrity. Some of those are due to training and some are inherent in your personalities. These are some of the best skills that any leader – especially technical leaders – need to have. Your foundational skills will help you develop skills like strategic thinking, building trust, networking and operational understanding. Some of the other critical skills you need to strengthen and learn include communication, business acumen, change management, and influence. Again, all these skills can be learned.

So, what can you do to more quickly develop your leadership skills? Here are a few things:

-Commit to your own leadership study

- Take instructional training
- Study different dimensions of leadership on your own
- Reflect on and learn from your own experiences

-Leverage the experiences & knowledge of others

- Identify and leverage mentors
- Form small discussion groups with peers

The right instructional training is one place to start. The ASA has developed courses that provide insights on leadership basics, influence, business acumen, cultural competency, and professional presence. Here are some upcoming opportunities for leadership training:

-Leading with Professional Presence in spring, 2020 at ASA headquarters in Alexandria (Final dates TBD)

-Preparing Statisticians for Leadership **i**n August at JSM 2020 in Philadelphia

-Leadership for Statisticians: The Bridge from Innovation to Practice **(tentative)** in September at RISW 2020 in Washington, DC.

I've had the pleasure of working with some statisticians who are outstanding leaders and have developed many of the skills I've listed above. Some have learned the skills through experience, and some have learned them thru a combination of experience and their own study. For me, leadership training and study has made a huge difference. I took my first, true leadership training course almost 20 years into my career. It opened my eyes to the power of effective leadership and the importance of leadership study. Since then I have been both a student and teacher of leadership for statisticians. Improving your leadership skills will not only improve your effectiveness and influence but will open up more opportunities for you to contribute in ways you may not have thought possible. So, the next time you walk out of a meeting frustrated because you couldn't get support for a great idea, think about investing in your leadership.

One more truth … It's never too late and never too early to become a "student of leadership." ■

SUMMARY OF A SURVEY ON DO'S AND DON'TS WHEN SUBMITTING A STATISTICAL MANUSCRIPT

Ilya Lipkovich, Eli Lilly and Company and Alex Dmitrienko, Mediana Inc

Recently BIOP conducted a survey of the editors and associate editors of several applied statistical journals with biopharmaceutical orientation (Statistics in Medicine, Statistics in Biopharmaceutical Research, Journal of Biopharmaceutical Statistics and selected editors from Pharmaceutical Statistics). The survey was designed to be open-ended and asked the editors to list, based on their experience, three Don'ts (oversights/weaknesses/red flags) commonly seen in submitted manuscripts that they advise authors to avoid. We also asked the editors to include three DO's (positive features or success factors). Of course, many DO's can be easily transformed into Don'ts and vice versa and the editors were encouraged to avoid listing trivial things that are well understood (such as plagiarism as a DO NOT) and focus on less obvious points that will be appreciated by statisticians who are preparing manuscripts for these journals. We received a total of 27 responses, each consisted of multiple DO's and Don'ts, most respondents provided 3 of each kind, some provided mostly Don'ts and some mostly DO's, perhaps depending on their general outlook (negative or positive).

Although the immediate motivation for conducting this survey was the desire to understand the reasons why so many of our own manuscripts have been rejected by these journals, we would like to share the results with the BIOP community as we strongly believe they will be useful to all authors of statistical manuscripts, whether novice or seasoned.

Because our sample size fell short of n=30, we refrain from any formal statistical inference in this article.

For convenience, we divided the areas covered by the respondents into 6 broad categories that are presented along with some detailed suggestions below. As both authors serve (or have served) as Associate Editors for Statistics in Medicine we felt free to add our own suggestions here and there.

Context of research and motivation

These include the need to provide proper motivation for research with such critical elements as *alignment with existing literature, novelty,* and *applicability*.

- DO include a convincing and easy to follow real-world example illustrating the problem and motivating your research.
- DO conduct a thorough and unbiased literature review ensuring good alignment with existing research. This not only means that the paper should present novel approach(es) but also helps avoid inventing new terminology and introducing new notation when established ones exist. As in almost any area of human endeavor, a good strategy is to stick to existing standards unless you can propose better ones.
- DO NOT include excessively long literature review, unless you are writing a review paper. Only cover literature relevant for your research.
- DO NOT try to solve artificial or non-existing problems. Always ask yourself: "is the method proposed applicable to a real-life problem?" Artificiality comes with many faces. For example, do not do research just to provide a Bayesian counterpart to a problem where an existing (frequentist) solution works well. Often authors "invent" methods involving multiple steps that somewhat arbitrarily combine existing procedures with little insight into why they should work better than available methods.
- DO NOT try to solve a special case when a more general problem has been already solved.
- DO NOT try to publish two very similar papers with a different order of authors.
- DO NOT write in the introduction that "unfortunately no approaches exist to handle this problem"; it is, in fact, quite fortunate for your research.

Structure and style of presentation

The key attributes are the *length, logical structure,* efficient use of *tables and figures*, use of *appendices* and *supplemental materials*. The recommendations are

- DO NOT write very long manuscripts (stressed by ¼ of responders!).
- DO make sure the structure of the paper is well thought out. As our respondents did not provide examples of poor structure, we would like to make a couple of specific suggestions:
  - o Although clinical journal standards require not to disclose results before the "results" section keeping the reader in suspense, this pedantic rule is not followed by many influential statisticians of our time who often present the summary of key findings in the introduction.
  - o Do not write a history of your research tracking how the ideas evolved in the course of writing this article, present the final view.
- DO present information efficiently, whenever possible graphical summaries are preferred over tabular summaries (in addition, tables can be moved to the appendix).
- DO NOT have an excessive number of tables and figures in the main text.
- DO provide detailed annotations for each figure which would allow the reader to understand the graph (and the context) even without reading the description in the text. This is also helpful for automatic generation of article summaries, as machine learning algorithms "like" to have figures explained by surrounding text.
- DO present proofs and other highly technical details in Appendices.
- DO make sure the paper is proofread by a native English speaker.

Simulation design

- DO make sure that simulations cover *relevant* cases.
- DO explain in nontechnical language why you choose particular scenarios.
- DO NOT choose only the scenarios that favor your method, in particular show how your method performs when the assumptions are not met.
- DO NOT choose as a comparator for your method a "strawman" (a pseudo-standard which is easy to beat); compare your method to a broad class of alternative approaches; in most settings there is no method that is uniformly better than all others, and it is important to identify the cases where alternative methods are superior to your method.
- DO NOT use simulations when you can make a point using an analytical argument.
- DO NOT present results with 9 decimal places, especially if you have run only 100 simulations per scenario.

Balancing theory with applications

- DO NOT overload the paper submitted to an applied statistical journal with mathematical equations and statistical jargon.
- The main results should be explained using language understandable by an intelligent non mathematically-oriented reader.
- DO provide intuition behind mathematical results.
- DO provide a real-life example (see also the section "context of research and motivation").

Proper framing your contribution

- DO NOT "oversell" and exaggerate the importance and novelty of your paper.
- DO NOT write your paper like a promotional dossier explaining how everyone is doing things wrong and "here we come and solve all the world's problems."
- DO provide a critical account of your research, stating gaps and limitations.
- Stating limitations of the proposed method is important; however, admitting your sins does not automatically mean forgiveness.

Reproducibility

- Describe methods in sufficient detail that can be reproduced.
- DO make the code and data sets available.

We would like to conclude this article with a DO (or rather a BE) suggested by one of our responders: "Be Brilliant!"

The authors are grateful to all editors who responded to the survey. ■

ASA BIOP NONCLINICAL BIOSTATISTICS WORKING GROUP SCIENTIFIC WORKING GROUP UPDATE

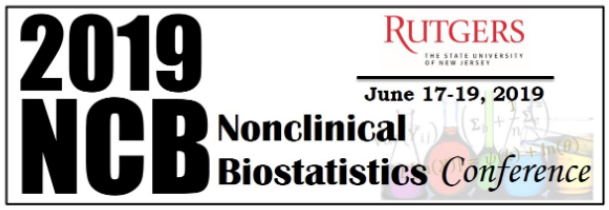
Steven Novick, AstraZeneca and Xin Huang, Abbvie

The Nonclinical Biostatistics Working Group (NCBWG), chaired by Steven Novick, oversees the biostatistics disciplines in all nonclinical activities of ASA-BIOP. Our goals, which are listed on our website page, include the promotion of biostatistics disciplines in nonclinical collaboration through scholarly activities, broadening the impact of biostatistical techniques in all areas of society, and promoting better understanding and interest by the general public in biostatistics beyond clinical applications.

In 2019, the 6th Nonclinical Biostatistics (NCB) Conference was held at Rutgers University from June 17 -19. One hundred and thirty attendees converged on this biennial conference for the presentation and discussion of scientific and statistical issues relevant to the field of nonclinical biostatistics. The program featured two short courses, 29 invited and contributed talks, 26 poster presenters, special sessions for graduate students, and was highlighted by keynote addresses, delivered by Dr. Karen Kafadar (ASA President-elect) and Dr. José Pinheiro (Global head of statistical modeling & methodology, Janssen). Also, at the NCB Conference, the NCBWG Best Nonclinical Paper committee awarded the "Stan Altan" best paper plaque to Lingmin Zeng for her paper on "General Framework for Equivalence Testing over a Range of Linear Outcomes with CMC Applications" (Zheng et. al., 2018). The student award committee also awarded two student participants the best-poster awards. The 2021 NCB Conference will be chaired by Xin Huang.

Rounding out NCBWG activities are two scientific working groups.

- The p-value group, headed by Stan Altan, starting with a round-table lunch at the NCB Conference and is now writing a manuscript to describe the current p-value controversy through the lens of practicing nonclinical biostatisticians.

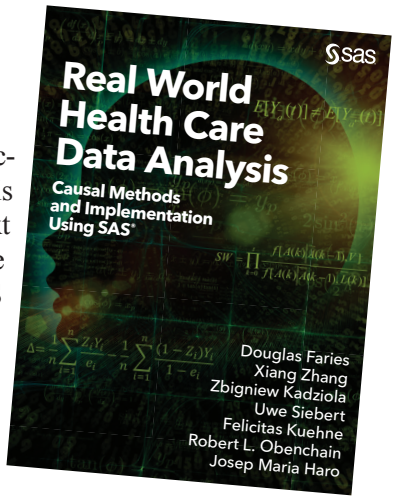


- The Nonclinical Bayes group, headed by Paul Faya and Perceval Sondag, is looking to improve Bayesian statistical method uptake within nonclinical biostatistics communities. In addition, the Bayesian group hopes to publish a manuscript on the use of informative priors for assay validation.

In all, we want to acknowledge all the committee members and participants in NCBWG and NCB conference for such a productive year. We look forward to a fruitful 2020! ■

BOOK: REAL WORLD HEALTH CARE DATA ANALYSIS: CAUSAL METHODS AND IMPLEMENTATION USING SAS

Authors: Douglas Faries, Xiang Zhang, Zbigniew Kadziola, Uwe Siebert, Felicitas Kuehne, Josep Maria Haro, Robert L Obenchain



Release: January 2020, SAS Press

Advances in communication and information technologies have led to an exponential increase in the collection of real-world data. Data in the health sector are not only generated during clinical research but also during many instances of the patient-clinician-payer relationship. This data serves as the basis for the growing use of real-world evidence (RWE) in medical decision-making. However, data itself is not evidence. A core element of producing RWE includes the use of designs and analytical methods that are both valid and appropriate for such data.

In 2010 we produced a book, Analysis of Observational HealthCare Data Using SAS (Faries et al. 2010), to bring together in a single place many of the best practices for real-world / observational data research. A focus of that effort was to make the implementation of best practice analyses feasible by providing SAS Code with example applications. However, since that time there have been improvements in analytic methods, coalescing of thoughts on best practices, and significant upgrades in SAS procedures targeted for real world research, such as the PSMATCH and CAUSALTRT procedures. In addition, the growing demand for real-world evidence and interest in improving the quality of real-world evidence to the level required for regulatory decision making has necessitated updating the prior work.

This new book has the same general objective as the 2010 text -- to bring together best practices in a single location and to provide SAS codes and examples to make quality analyses both easy and efficient. The main focus of this book is on causal inference methods to produce valid comparisons of outcomes between intervention groups using non-randomized data. Our goal is to provide a useful reference to help clinicians, epidemiologists, health outcome scientists, statisticians, data scientists etc. to turn real world data into credible and reliable real-world evidence.

The opening chapters of the book present an introduction of basic causal inference concepts and summarize the literature regarding best practices for comparative analysis of observational data. The next portion of the text provide detailed best practices, SAS code and examples for propensity score estimation and traditional propensity score-based methods of matching, stratification and weighting. In addition to standard implementation, we present recent upgrades including automated modeling methods for propensity score estimation, optimal and full optimal matching procedures, local control stratification, overlap weighting, new algorithms that generate weights that produce exact balance between groups on means and variances, methods that extend matching and weighting analyses to situations comparison more than 2 treatment groups, and a model averaging approach to let the data drive the selection of the best analysis for your specific scenario. Two chapters of the book focus on longitudinal observational data. This includes an application of marginal structural modeling to produce causal treatment effect estimates in longitudinal data with treatment switching and time varying confounding and a target trial replicates analysis to assess dynamic treatment regimes. In the final section of the book we present analyses for emerging topics: reweighting methods to generalize RCT evidence to real world populations, sensitivity analyses and best practice flowcharts to quantitatively assess the potential impact of unmeasured confounding, and an introduction to using real-world data and machine learning algorithms to identify treatment choices to optimize individual patient outcomes. ■