Paper AS07

Database Tools Needed for Personalized Study Participant Risk Surveillance

Martin Bauwens, MMS Holdings, Canton, MI, USA
Chris Hurley, MMS Holdings, Canton, MI, USA

ABSTRACT

Despite the most rigorous pre-clinical trial methodologies and safety protocols, human drug trials have a degree of risk to the participants. New tools can be constructed to foster trust and facilitate participant safety, employing variation within participant traits and identify risk. Here we will discuss opportunities to identify risk based upon control participant data across several prior studies. We will demonstrate that grouping control participants within demographic, health, geographic and other factors and then calculating normal, borderline and atypical ranges can define participant risk within specific populations. Specifically, by knowing these ranges in the context of equivalent peers, a more individualized risk appraisal can be sent to safety monitors and investigators as a mechanism of caution. By developing of an easy-to-use tool that specifies risk in a more person-centric way, trust, safety and confidence will be enhanced in clinical trials.

INTRODUCTION

This paper is not a how-to, or step-by-step instruction manual, for developing risk analysis and mitigation strategies with regards to your clinical trials data. It does however present a very basic introduction to the use of various statistical or analytical tools that may be helpful when trying to uncover any quality or safety concerns. It is said that well-controlled randomized clinical trials are the gold standard but variability in trial design, data sources, patient characteristics, disease features and even how data may be collected might necessitate a close examination of the relationships and quality of the data.

DATA NEEDS AND ORIGINS

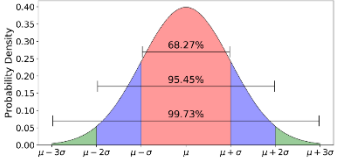
The typical clinical trial protocol is generally loaded with opportunities for data to support and prove the safety and efficacy of the device or drug intervention under study. References to summaries of existing (non)clinical studies in the program may lead to potential data that can be utilized to assess current study data. Primary, secondary, and exploratory objectives and other endpoints may facilitate the analysis of real-world data. Data from electronic patient reported outcomes may be used to support efficacy or safety analyses. Study stopping rules may employ external vendor data within lab or other findings domains. A look at the typical study schedule will reveal data that may have relationships across visits or domains while posing challenges to cleanliness and accuracy. We must ensure the data is clean and appropriate for the endpoints of the study and more importantly, the safety of each study participant.

IS THAT NORMAL?

Several factors such as age, gender, race, medications, and others may determine what is normal in regard to laboratory test values. Normal, borderline and atypical values are generally derived from data from thousands and thousands of people for the particular lab tests and analyzers evaluating them. By understanding the general shape of the clinical study data in context with these normal ranges it is possible to find the data values that are either clinically significant or have been entered or converted in some way that are way out of line with expectations.

Density plots, like the example below can reveal the shape of your data. Any values within the green tails may trigger actions such as communication internally to a data management team or externally to investigators and site staff. Known atypical lab values may be considered in decisions to:

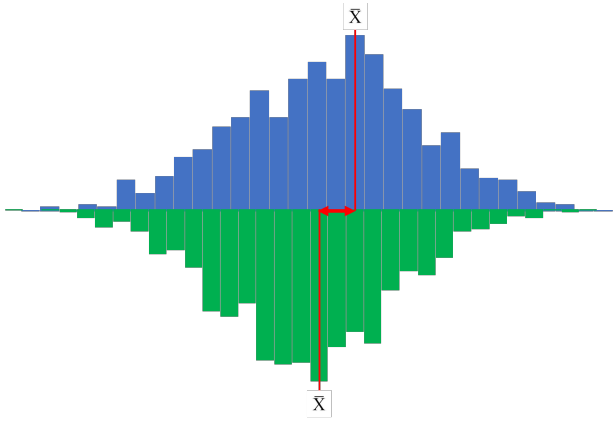
- Keep the participant in the study
- Remove the participant from the study
- Advise the participant to seek medical care
- Provide clinical/medical care to the participant



WHAT'S SHAPE GOT TO DO WITH IT?

Generally you need to understand the shape of your data to understand how to analyze it. Whether data is normal or non-normal will determine if you use parametric or non-perimetric analysis, but this is beyond the scope of this paper

though some considerations for analysis will be presented in a section below. In general your lab data should be normally distributed yet some group traits may result in a very different curve than the broader population normal curves. Phenotypic, geographic, SES, medical history, current medications and supplements influence many baseline physiological processes. Study participants with certain identifiable traits can have lab values distributed much different than the broader population. Specific harm events can be signaled with less latency and greater validity when normalized distributions consider within groups effects. Genotypic and phenotypic grouping may also have group effects. Some literature has been included as recommended reading if you want to learn more about within groups distributions outside full population distributions.

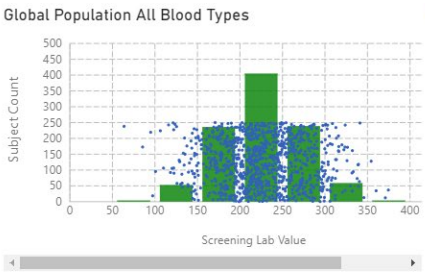


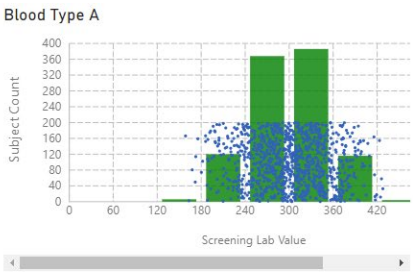
WELL STRUCTURED DATA ENABLES CROSS STUDY COMPARISON

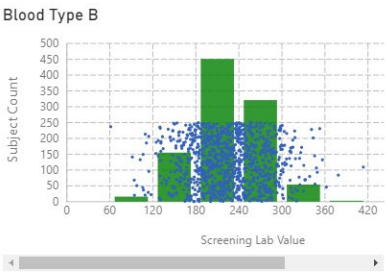
One of the best aspects about data standardization is the ability to pool data quickly and easily. The Clinical Data Standards Consortium (CDISC) has a model called the Clinical Data Acquisition Standards Harmonization Model (CDASH) which provides direction on how data should be collected and databased. The Study Data Tabulation Model (SDTM) is the next level of standardization and works best when it is fed CDASH data, but other data can be mapped to this model. This standardization is closely aligned with the collected data yet specifies several data structures of various types of data such as events, interventions, findings, trial design and special purpose datasets. The Analysis Data Model (ADaM) is the final level of standardization where SDTM data is transformed and optimized for analysis purposes.

STACKED DATA

The SDTM is particularly good for logical grouping of data across studies. Though the SDTM standard isn't exact, after tweaking for variable lengths and any other small variance across study datasets great utility can be had when pooling SDTM datasets. The histograms and whisker plot examples below show some example lab data values taken at screening but could, in theory, be collected at any timepoint. By grouping data for an entire population then comparing it to other groups of data certain insights are almost instantly recognized. For instance in the plots below, we see that those with Blood Type A may have a signal should the screening lab value of interest be less than 200 while the same results in the Global grouping and the Blood Type B grouping would be quite normal. Visualization tools may be set up to produce these insights using CDASH/SDTM well before study statistical analysis plans and ADaM data are ready from programming and can give an early look at safety signals within the collected data.



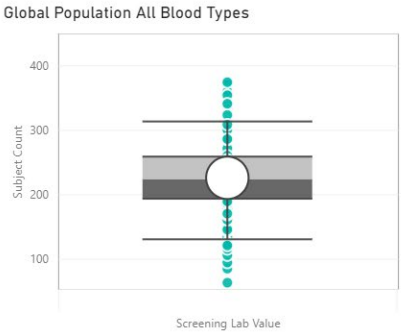


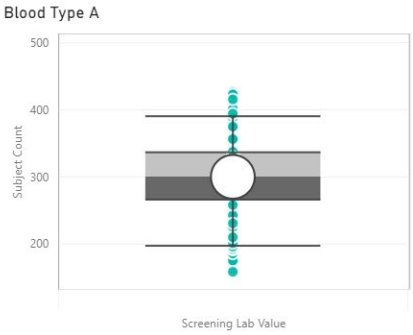


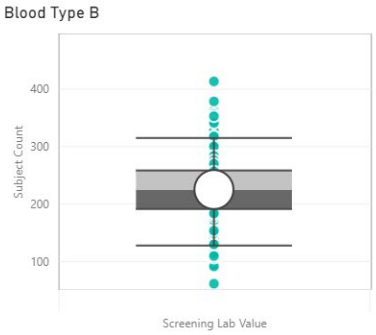
Values less than 131 are statistically defined as outliers within the normal curve and may indicate further investigation or trigger an alert.

Values less than 200 are statistically defined as outliers within the normal curve and may indicate further investigation or trigger an alert.

Values less than 128 are statistically defined as outliers within the normal curve and may indicate further investigation or trigger an alert.







PARAMETRIC OR NON-PARAMETRIC?

As with all things that can be described with numbers there are certain limits and caveats that must be considered if the application of those numbers is to be meaningful. Often the most critical question that should be asked of data to be used in a statistical framework should be, is the data parametric or not? If the data are parametric, it can be expected that those data will follow normal distribution with the mean being the center point, a standard deviation explaining 86 percent of the data, at two standard deviations 95% of the data are explained and so on. In a non-parametric scenario these segments are not defined in such a predictable way. Parametric vs. non-parametric characteristics of data are quickly assessed by comparing the mean to the median in conjunction with calculation of the kurtosis of the data.

Creation of calculated values table

There are a number of ways to proceed with defining and then proceeding with data that is assessed on it's parametric nature or lack thereof. For this paper it is suggested that data are processed in steps to accommodate the nature of those data. Here it is suggested that upon grouping data with respective kurtosis calculated. Those results will then be stored in a temp table. Cases with > 31 cases and a kurtosis between -1 and +1 would be selected for the next stage of processing where the mean and standard deviation are calculated and a flag for the type of calculation be set parametric. These data would populate a second temp table. A third temp table where the number of cases was > 31 and thee kurtosis was < -1 and > +1 would have the median, interquartile range(IQR) and IQR * 1.4 defined and a flag set to non-parametric. Cases remaining that are <= 31 would be reapplied with other like groups to meet sufficient power for application(with temp tables 1 and 2 recreated to accommodate the small samples in other groupings). Temp tables one and two would now have their labels changed and be applied to an integrated dataset using a UNION command.

Resulting Table headers:

Grouping	Cases	-2 SD/IQR * -1.4	-1 SD/IQR Floor	Central Tendency	+1 SD/IQR Ceiling	+2 SD/IQR * +1.4	Parametric/Non-Parametric

Following the creating of this table any values that are received can be queried against this table and the values compared.

PUTTING IT ALL TOGETHER

Various observations show that treating lab values as homogeneous when defining normal ranges may miss signals of risk in some populations. Academic research has shown that demographic and phenotypic traits can be used to create specific normalized ranges that are exclusive to traits or groups of traits within individuals. Use of more specific normalization curves that reflect observable person traits can add an additional dimension of awareness of safety and efficacy signals in research participant data. These signals can provide risk and success indicators not seen when using homogenous data that does not consider specific traits and groups of traits.

The data used to create normal ranges for lab values that are reflective of within person groups is available though application of data from prior studies. Clinical data in CDISC format will maintain the same structure and data types across studies. For this reason CDISC data can be stacked with one study atop of another across all available repositories. By limiting the data to control participants the obvious confound of study drugs is removed. This broad pool of study data will be very large and provide good power across groups found in data. These now combined data can be used to define normal and atypical ranges on values in ways that consider within person traits. This entire process can be done relatively easily using simple database commands that only take moments to program.

Ranges by trait and groups of traits can now be stored and used against new data arriving from sites and sponsors. By employing computed ranges from your own big data and the groups within them, signals that are specific to the traits a person has can be communicated to study staff and primary investigators. These reports can be completely automated, and email can be generated using SQL Server Integration Services or other packages to immediately flag signals for the research staff and primary investigator to assess. Ultimately, by applying historical data to identify trait-based signals within lab data, safety and efficacy surveillance is enhanced. This enhancement is the product of considering how traits can influence normal ranges on lab values ad then applying those specific signals that would be lost if looking at the data in a homogeneous way.

APPLICATION OF THESE DATA

Results from calculated ranges based on traits and groups of traits can be immediately applied to any incoming data. By joining a table or tables with more specific ranges on lab tests with incoming data in an Extract Transform and Load (ETL) job individual cases can be flagged. Flagged cases, the recorded lab value and the normal range for her/his trait-based range can then be flagged in the database and or sent to the study staff and primary investigator. By being provided with these flagged data those sent the findings can appraise the results and consider if the person in question should be withdrawn, receive treatment or other risk mitigation strategies. At the same time, these signals could provide much faster feedback on of certain traits enhance or reduce efficacy. In both scenarios, use of homogenous full population ranges may miss this specificity. This specificity is expected to reduce the risk in clinical trials as it is a well-tuned warning flag that has not been applied well across the industry yet.

RECOMMENDED READING

Recommended reading for those interested in learning more.

- Kuś, Aleksander et al. “Variation in Normal Range Thyroid Function Affects Serum Cholesterol Levels, Blood Pressure, and Type 2 Diabetes Risk: A Mendelian Randomization Study.” Thyroid : official journal of the American Thyroid Association vol. 31,5 (2021
- Chang, Sheng-Nan et al. “Sex, racial differences and healthy aging in normative reference ranges on diastolic function in Ethnic Asians: 2016 ASE guideline revisited.” Journal of the Formosan Medical Association
- Varma, Adarsh et al. “African Americans Demonstrate Significantly Lower Serum Alanine Aminotransferase Compared to Non-African Americans.” Journal of racial and ethnic health disparities vol. 8,6 (2021): 1533-1538.
- Cogswell, Mary E et al. “Estimated 24-Hour Urinary Sodium and Potassium Excretion in US Adults.” JAMA vol. 319,12 (2018): 1209-1220.

CONTACT INFORMATION (HEADER 1)

Your comments and questions are valued and encouraged. Contact the author at:

Author Name: Martin Bauwens
Company: MMS Holdings Inc.
Address: 6880 Commerce Blvd.
City / Postcode: Canton, MI 48187
Work Phone: +1(226) 235-1249
Email: mbauens@mmsholdings.com

Author Name: Chris Hurley
Company: MMS Holdings Inc.
Address: 6880 Commerce Blvd.
City / Postcode: Canton, MI 48187
Work Phone: +1(734) 245-0469
Email: mbauens@mmsholdings.com

Web: mmsholdings.com