PhUSE US Connect 2019

Paper DV03

Interactive Visualization of SEND Clinical Pathology Data Using R Shiny

Kevin Snyder, Ph.D., U.S. Food and Drug Administration, Silver Spring, MD, USA

ABSTRACT

The standardization of nonclinical study data by CDISC-SEND has created the opportunity for the development of open source data analysis software that can be collaboratively developed by both the pharmaceutical industry and government regulators through consortia such as PhUSE. R Shiny® facilitates the development of user-friendly, web-based "Apps", providing a platform for the analysis and visualization of SEND datasets. Clinical pathology data from nonclinical toxicology studies can be difficult to digest when presented as group means in data tables, as is common practice in study reports, due to the large number of potentially correlated analytes collected across treatment groups, sexes, and potentially multiple timepoints. An R Shiny application has been developed to allow end users to comprehensively examine these datasets, using a variety of analytical and visualization methods, with relative ease. The source code for this application is publicly available on the PhUSE GitHub website: https://github.com/phuse-org.

INTRODUCTION

In December 2014, the United States Food and Drug Administration (FDA) finalized a binding guidance requiring specific types of clinical and nonclinical study reports to be submitted with standardized electronic data. A list of data exchange standards currently supported by the FDA can be found in the FDA Data Standards Catalog, along with requirement dates for each. Any study that can be supported by a data exchange standard and was initiated after that standard's requirement date is required to be submitted with standardized electronic data. The only data exchange standard currently supported for nonclinical study data is the Standard for Exchange of Nonclinical Data (SEND), which was created by the Clinical Data Interchange Standards Consortium (CDISC). The requirement dates for SEND Version 3.0 are December 17, 2016 for study reports included in New Drug Applications (NDAs) and December 17, 2017 for study reports included in Investigational New Drug Applications (INDs), and the nonclinical studies supported by SEND Version 3.0 include single-dose general toxicology, repeat-dose general toxicology, and carcinogenicity studies. This means that for new IND submissions, any single-dose general toxicology, repeat-dose toxicology, or carcinogenicity study initiated after December 17, 2017 will need to be submitted with SEND datasets. Clinical pathology assessments are conducted for all single-dose general toxicology, repeat-dose toxicology, and carcinogenicity studies and generate relatively large and complex data that has historically been submitted to regulators in the format of static tables. The standardization of these data via SEND has provided the opportunity for the collaborative development of data visualization software that may be useful for any organization reviewing SEND-formatted study data.

IMPACT OF SEND ON CLINICAL PATHOLOGY DATA VISUALIZATION

The traditional toxicology study reports included in regulatory submissions describe the clinical pathology results via narrative text and summary tables. The summary tables generally report the mean, standard error, and number of subjects for each test within each treatment group along with the results of a statistical analysis of the group means compared to the control group, e.g. Dunnett's Test. Additional tables are also included to record the test results in each individual subject, making extraction of this data from the study report documents for the purpose of data visualization technically feasible but terribly inefficient.

Adoption of the SEND standard has created the opportunity for toxicology study data generated by disparate organizations to be visualized using the same software. In a SEND dataset, the results of clinical pathology evaluations can be found in the Laboratory Test Results (LB) domain. Other SEND domains, e.g. Demographics (DM) and Trial Sets (TX), store information about the study subjects and their treatment allocations. The standardized formatting of this information allows computer programs to reliably produce visualizations of the results of any toxicology study encoded in SEND format.

DATA VISUALIZATION METHODOLOGIES WITH RELEVANCE TO CLINICAL PATHOLOGY DATA

When generating figures for the visualization of data choices must be made about how much and which data to include in each single figure. Data may be filtered on basis of several factors including: test, sex, treatment, and day of test. In addition, it may be useful to transform the data, e.g. z-score or percent change from control, prior to visualization to accommodate the visualization of several test results with vastly different numerical result ranges. It may also be useful to represent the data in terms of change from baseline if tests were recorded on multiple days in each subject, and then these values could still be subsequently transformed as z-scores or percent change from baseline if desired. After the data has been properly filtered and/or transformed it can be visualized via several methods that can generally be categorized as plotting group summary data or individual subject data.

METHODS OF PLOTTING OF GROUP SUMMARY DATA

Figures displaying group summary data typically represent group means with bars, points, or lines and often use error bars to represent group variance, usually as the standard error of the mean. Bar plots are familiar to most audiences and can be useful for display treatment effects on multiple parameters in the same figure. Point plots convey the exact same information as bar graphs but can more easily display an additional dimension, e.g. sex, via the shape of the points; however, they are slightly less familiar to some audiences. Line plots are essentially the same as point plots with the addition of lines connecting the points, which can be used to help viewers track changes in parameter values across the x-axis, e.g. across time. Box and whisker plots are another way of presenting group summary data by displaying the median and quartile ranges, allowing viewers to assess the skewness of the distributions rather than assuming symmetrical gaussian distributions as is implied by plots of mean and variance.

METHODS OF PLOTTING OF INDIVIDUAL SUBJECT DATA

Sometimes it can be more useful to display individual subject data than summary statistics, for example if you are interested in the responses of a particular set of subjects within a group. Line graphs can be particularly useful when tracking changes in individual subjects over time or across tests, e.g. to check if the same subject has the highest value for a set of tests. Box and whisker plots can also be modified to show individual data points in addition to the quartile summary statistics. Another advantage of visualizing individual subject data is that it allows viewers to visually assess the correlations among test results across subjects, typically via a scatter plot matrix. Further multivariate analyses, taking into account the correlations of test results across subjects, such as principal component analysis, can be performed to evaluate the response of each subject for a set of parameters and look for clustering of treatment groups. This analysis can be particularly useful when treatments produce a consistent pattern of effects across tests within subjects and can be used to convey a large amount of information in a single figure.

CLINICAL PATHOLOGY DATA VISUALIZATION USING R SHINY

In collaboration with the PhUSE Nonclinical Scripts Working Group, an R Shiny web application was developed to showcase methodologies that can be used to visualize clinical pathology results from SEND-formatted study data. The application reads data from publicly available mock SEND datasets and can be freely accessed at:

https://phuse-nonclinical-scripts.shinyapps.io/LBapp/

In addition, the source code for the R Shiny application is available on the PhUSE GitHub website at:

https://github.com/phuse-org/phuse-scripts/tree/master/contributed/Nonclinical/R/LBapp/

DEMONSTRATION OF FUNCTIONALITY OF R SHINY CLINICAL PATHOLOGY DATA VISUALIZATION APPLICATION

The application provides a web-based HTML/CSS/JavaScript-based user interface while performing all computation and producing all graphics using R. Figure 1 shows the layout with the user interface options on the left and the visualization output on the right, in this case a group summary data table. The three open menus displayed are: Tables, Figures, and Select Dataset. The Table and Figures menus provide options for how to visualize the data as selected, filtered, and transformed based on the options in the menus below. The Select Dataset menu allows the user to load datasets from PhUSE GitHub site, in this case the publicly available PDS-2014 mock SEND dataset (http://senddataset.org/) has been selected. The Select Tests menu allows the user to filter tests by category (LBCAT) and then select tests of interest, using the LBTESTCD to identify each test (Figure 2A). If the same LBTESTCD is used in more than one test category, then the first four letters of the test category are appended to the beginning of the test name, e.g. CLIN_PROT and URIN_PROT. Users are able to select tests of interest and interactively select the order that the tests will be displayed in tables and figures by clicking and dragging the test name icons. The settings menu allows users to filter and transform data to fit their data visualization needs (Figure 2B). Users can choose to display groups across all treatments and days or display groups across treatments by selected day or across days by selected treatment. Data can also be filtered by sex to display groups within one or both sexes. The last two options allow the user to transform the data by representing it as change from baseline, in the case where data from the same subjects is available across multiple days, or by representing it as z-scores or percent change from control or baseline.

Figure 1: R Shiny Application User Interface with Group Summary Data Table Displayed

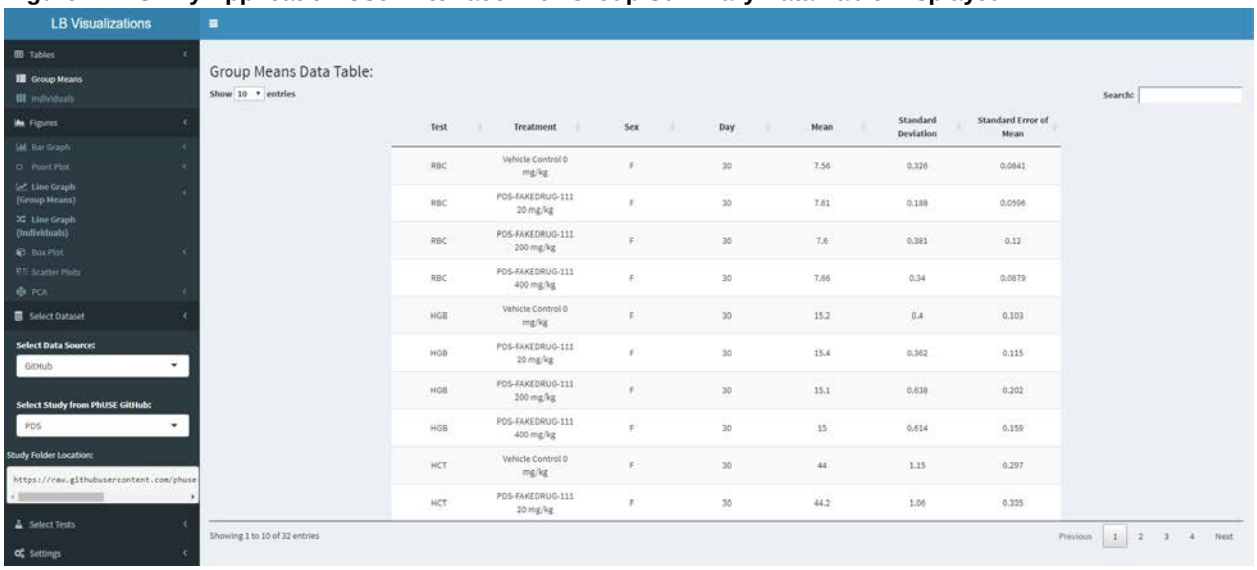
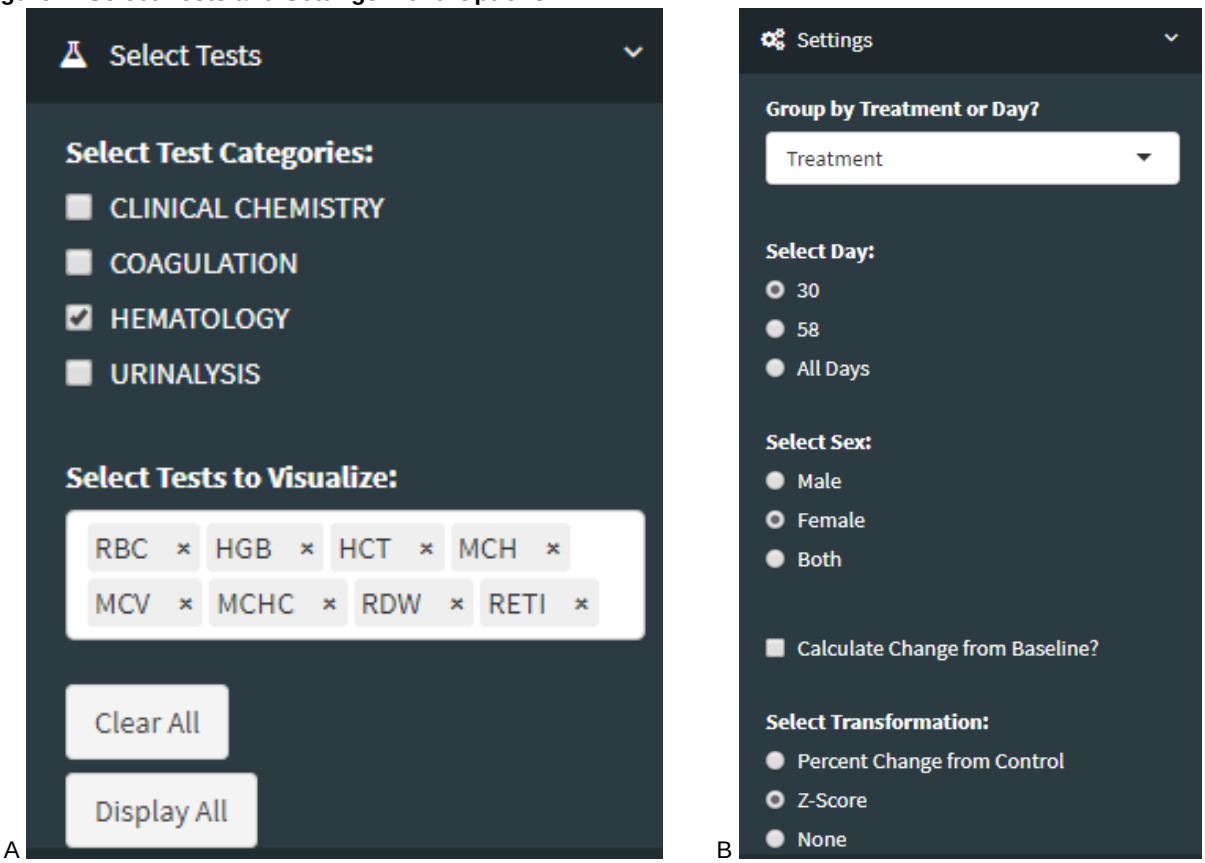


Figure 2: Select Tests and Settings Menu Options



The following examples demonstrate how the same data, i.e. red blood cell parameters, can be viewed from several different perspectives using the data filtering, transformations, and visualization methodologies available in the R Shiny application. Notably, the magnitude of percent change in the red cell distribution width (RDW) and reticulocyte count (RETI) in females is highlighted with respect to the other parameters the bar chart (Figure 3) and point plot (Figure 4) whereas the dose-dependent decreases in mean corpuscular hemoglobin (MCH) and mean corpuscular red cell volume (MCV) are more visible in the z-scored data presented in the line graph (Figure 6) and box plot (Figure 8). The line graph in Figure 5 provides a clear demonstration that the pronounced increase in reticulocyte

count observed at the high dose on study day 30 was not present after the recovery period on day 58. The correlations among tests that can be clearly observed in the scatter plot matrix (Figure 9) are also evident in consistency of z-scored individual subject values across correlated tests in line plot of individual subjects results across parameters for the vehicle and high dose treatment groups (Figure 7). The principal component analysis (Figure 10) demonstrates that the most prominent treatment effect was observed in the highly correlated RDW and RETI test results as the axis of separation of treatment groups is clearly pulled in the shared direction of the loading vectors for these two tests. The lack of treatment effect or correlation with other parameters observed in red blood cell count (RBC) and mean corpuscular hemoglobin concentration (MCHC) are evident by the relative isolation of their loading vectors and their perpendicular direction with respect to the axis treatment group separation. The pairing of the loading vectors for MCH and MCV as well as for hemoglobin (HGB) and hematocrit (HCT) indicate strong correlations between these pairs of tests and their directions indicate treatment-related decreases in both parameters, although not necessarily in the same subjects across these pairs.

Figure 3: Bar Graph of Percent Change in Red Blood Cell Parameters from Females on Day 30

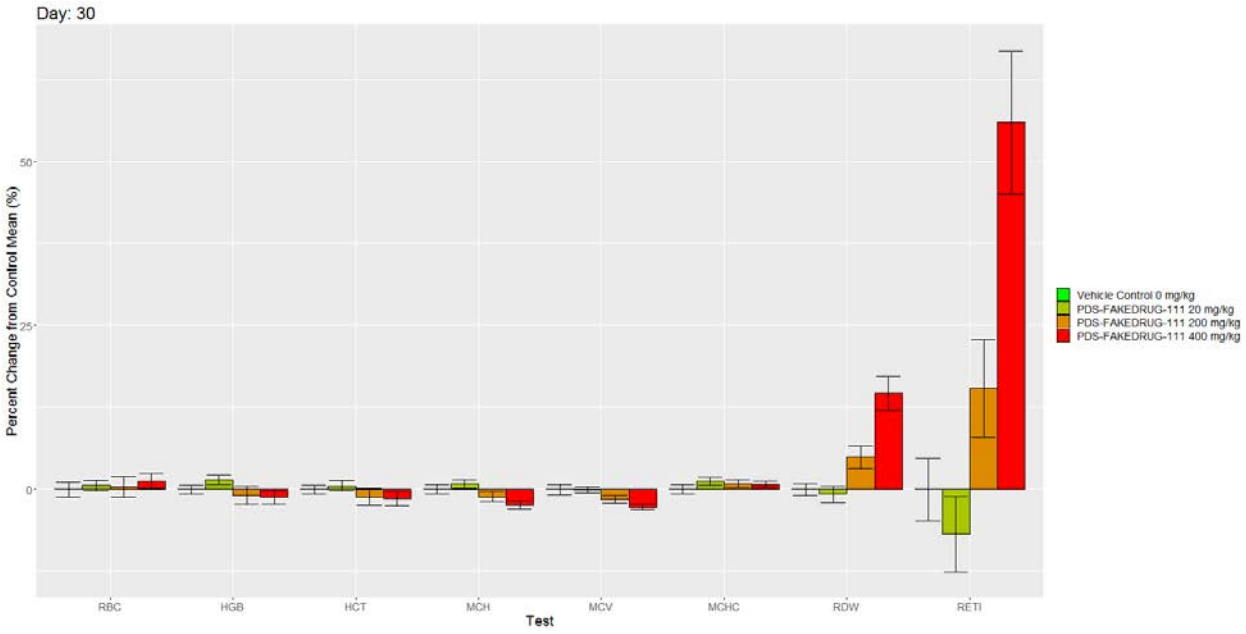


Figure 4: Point Plot of Percent Change in Red Blood Cell Parameters from Males and Females on Day 30

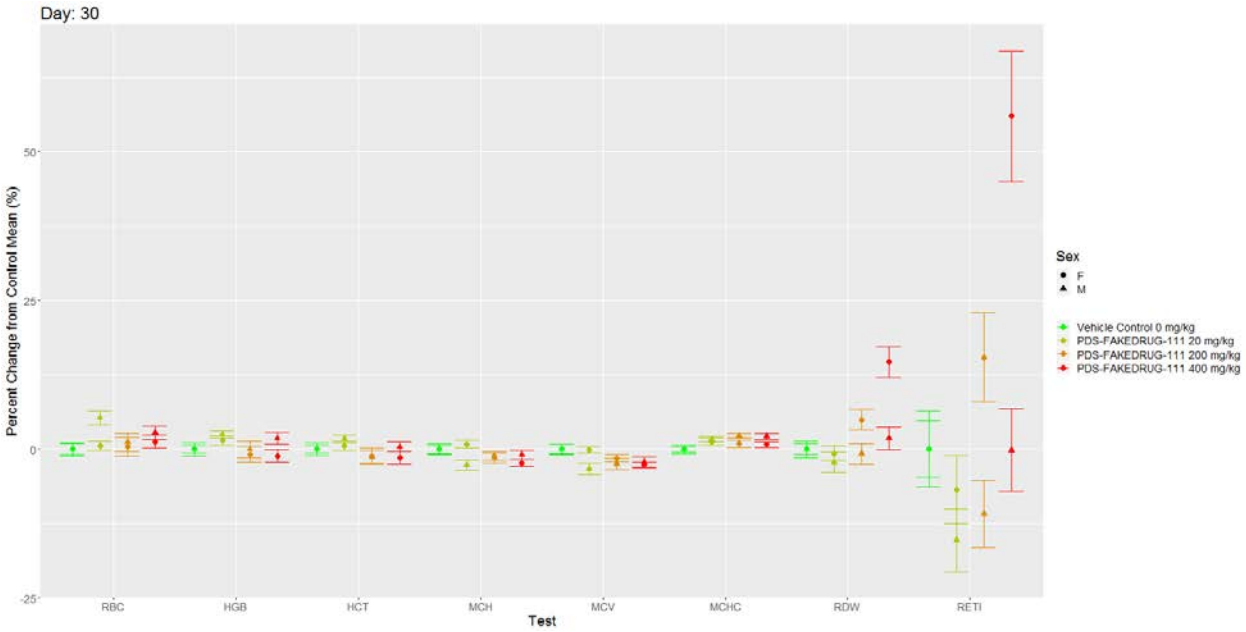


Figure 5: Line Graph of Percent Change in Reticulocyte Count from Females on Day 30 and 58

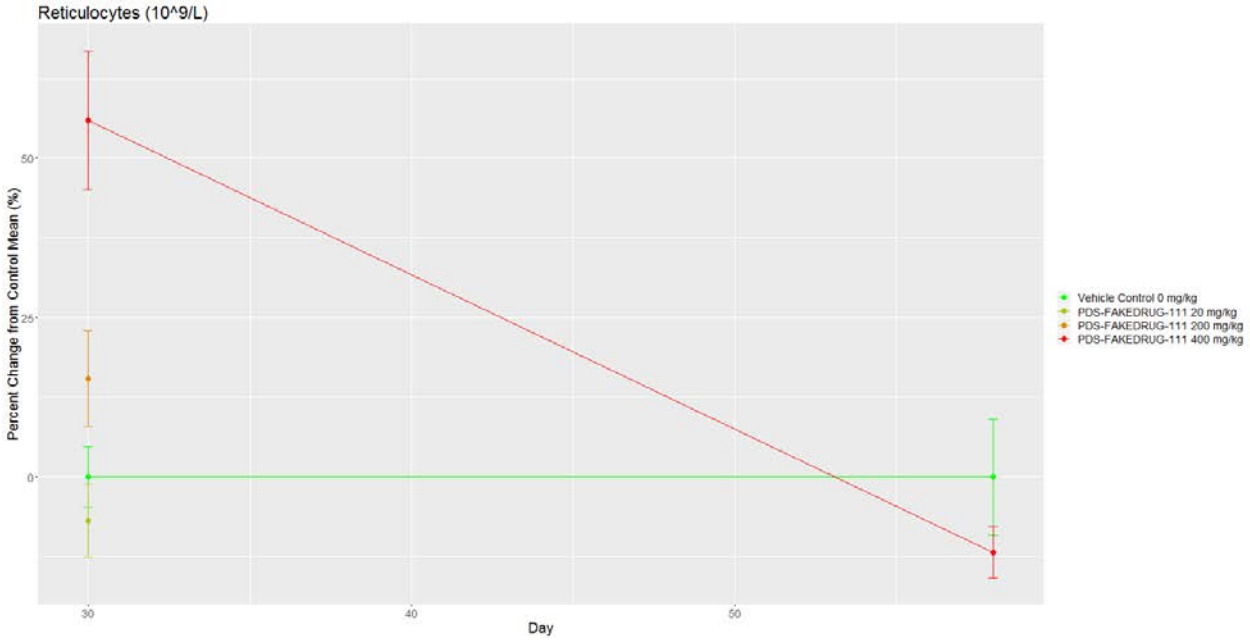


Figure 6: Line Graph of Z-Scored Red Blood Cell Parameters from Females on Day 30

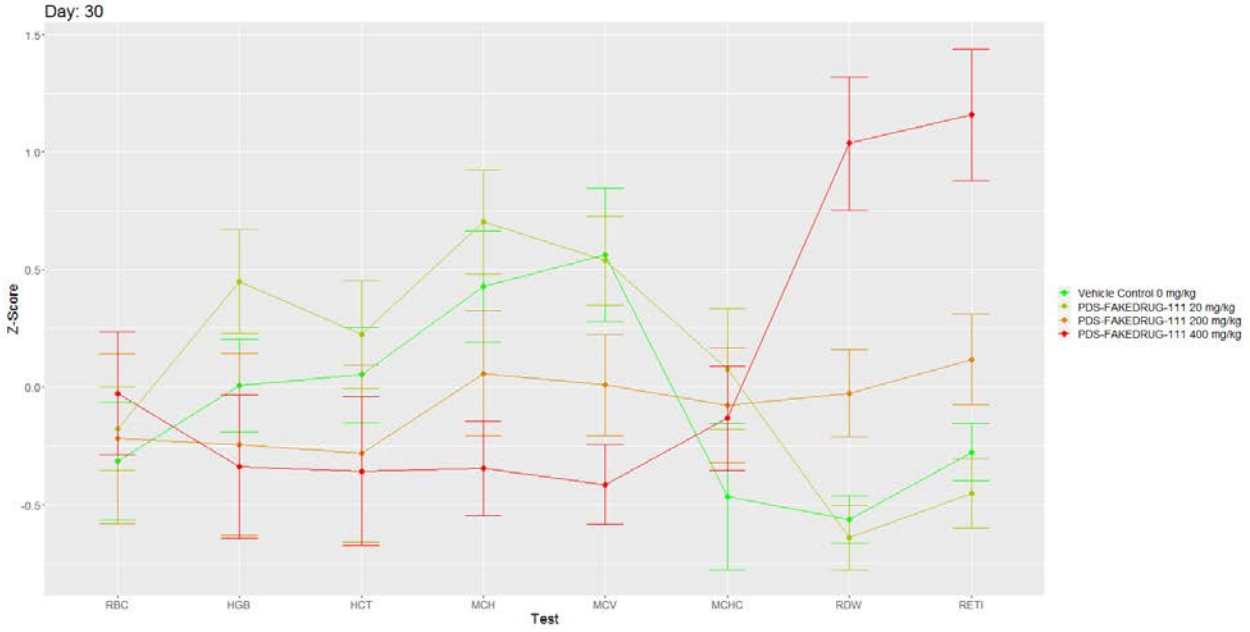


Figure 7: Line Graph of Z-Scored Red Blood Cell Parameters from Individual Females on Day 30

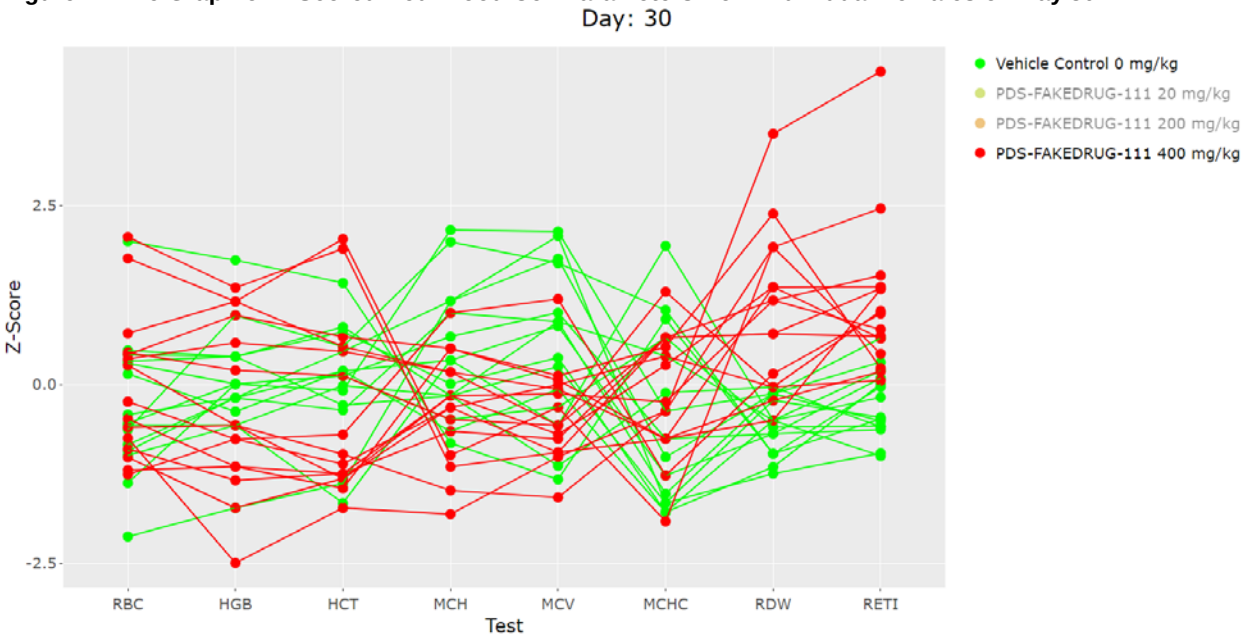


Figure 8: Box Plot of Z-Scored Red Blood Cell Parameters from Females on Day 30

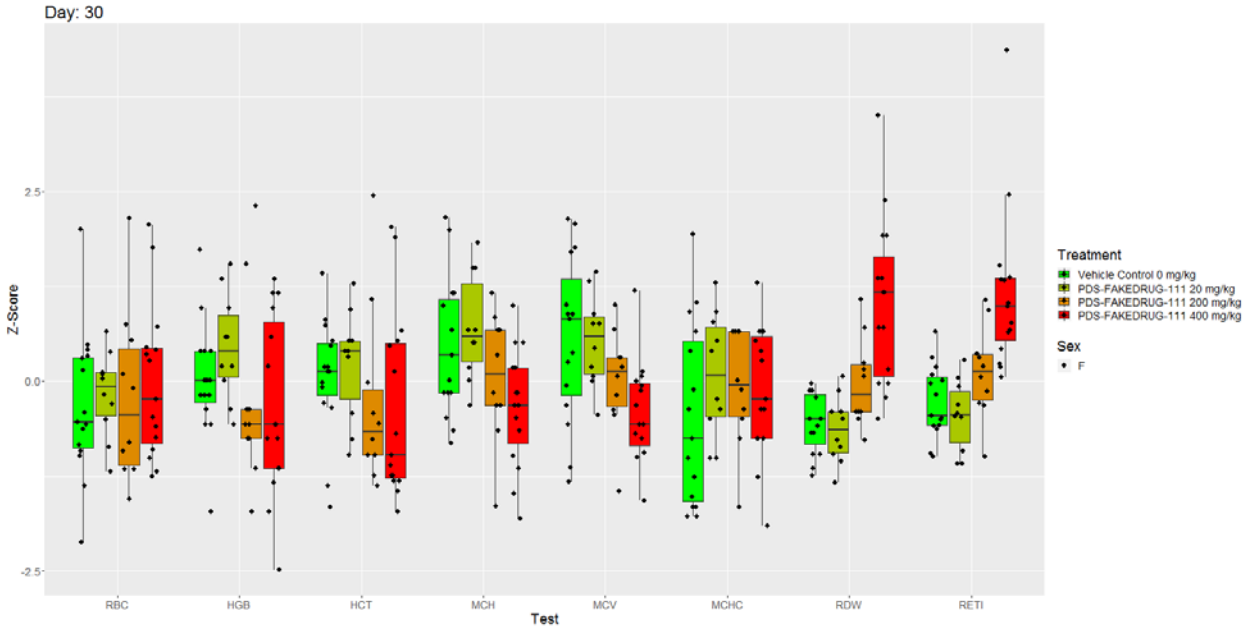


Figure 9: Scatter Plot Matrix of Red Blood Cell Parameters from Females on Day 30

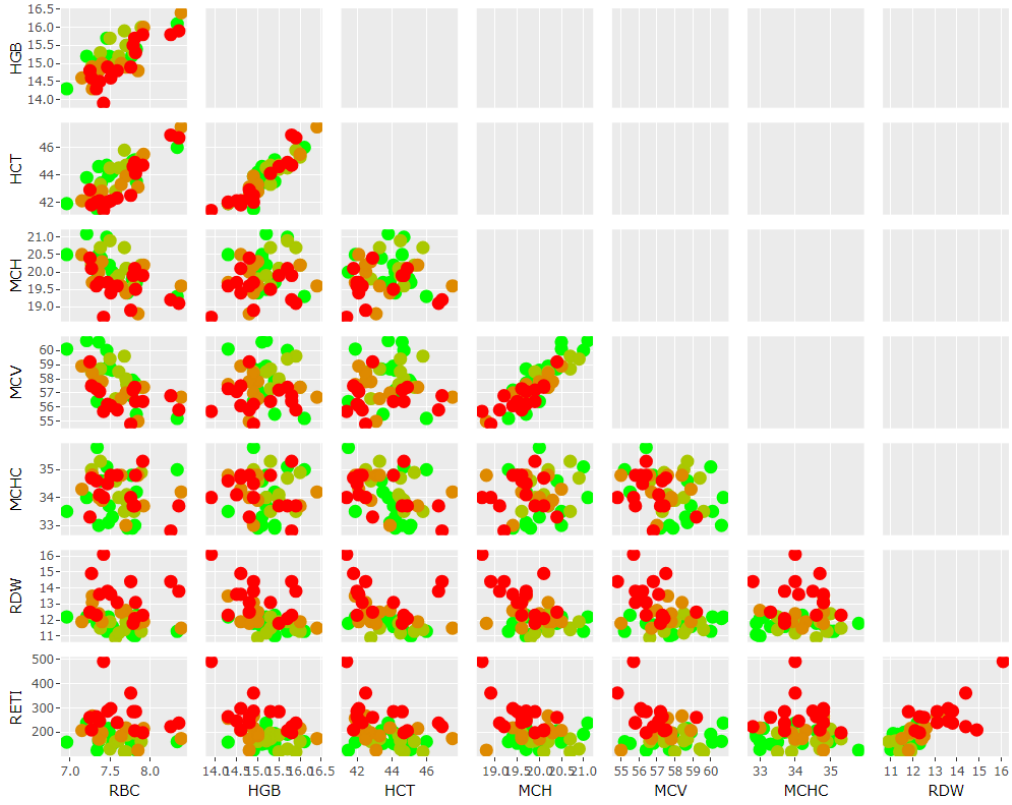
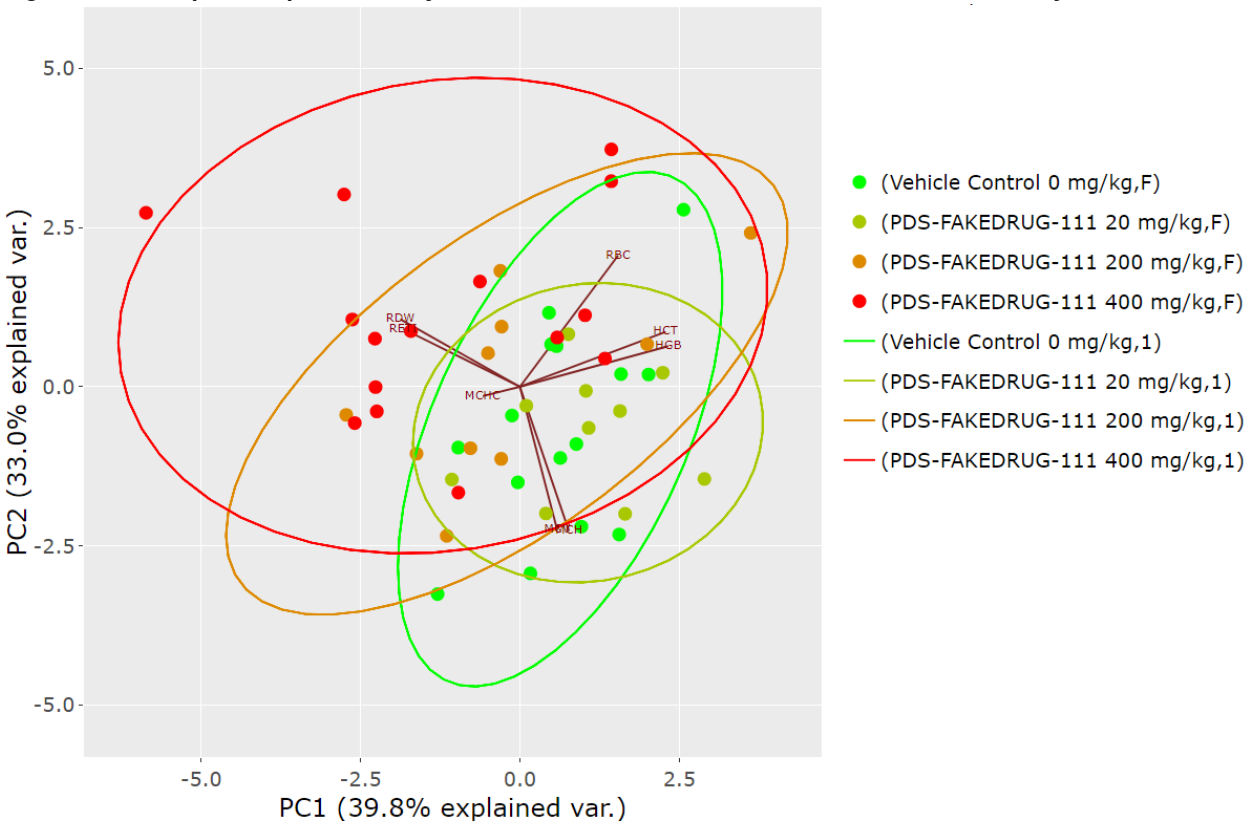


Figure 10: Principal Component Analysis of Red Blood Cell Parameters from Females on Day 30



R SHINY

R is a free and open source software environment for statistical computing and graphics that is publicly available at https://www.r-project.org. Shiny is an R package developed by RStudio® to enable R programmers to easily build interactive web applications without requiring R developers to know or write any HTML, CSS, or javascript code, although it does include functions that allow advanced users to embellish their web applications with custom HTML, CSS, and/or javascript code. The R Shiny web application is currently hosted on a public R Shiny server provided by RStudio.

CONCLUSION

The standardization of nonclinical study data via SEND has facilitated the collaborative development of data visualization software that can be used to greatly enhance the ease of interpretation of study results, particularly for study components that generate myriad potentially correlated numeric results such as clinical pathology assessments. Many opportunities for future development of additional visualization methods remain open within the context of clinical pathology data, e.g. the LB domain in SEND, as well as for other SEND domains and for cross-domain analyses. The PhUSE Nonclinical Scripts Working Group will continue to provide resources to the SEND user community by continuing to develop and share new data visualization methods and open source software implementations of these methods.

DISCLAIMER

This publication reflects the views of the author and should not be construed to represent FDA's views or policies.

ACKNOWLEDGEMENTS

The author would like to acknowledge members of the PhUSE Nonclinical Scripts Working Group, particularly Bill Houser, Hanming Tu, and Bob Friedman, for their assistance in the creation of the R Shiny web application.

CONTACT INFORMATION

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

Author Name: Kevin Snyder, Ph.D.
Company: United States Food and Drug Administration
Address: 10903 New Hampshire Avenue
City / Postcode: Silver Spring, MD 20993
Work Phone: 301-796-7252
Email: kevin.snyder@fda.hhs.gov