Cross-Study Analysis of SEND Datasets Using an R Package: sendigR

Kevin Snyder[1], Yousuf Ali[2], Jesse Anderson[1], William Houser[3], Daniel Russo[4], Bo Larsen[5], Mark Carfagna[6]
[1]United States Food and Drug Administration, Center for Drug Evaluation and Research, White Oak, MD, US; [2]Oak Ridge Institute for Science and Education, Oak Ridge, TN, US;
[3]Bristol Myers Squibb, New Brunswick, NJ, US; [4]Rutgers University, New Jersey, US; [5]Novo Nordisk, Copenhagen, Denmark; [6]Eli Lilly, Indianapolis, IN, US

Introduction

The Standard for Exchange of Nonclinical Data (SEND) developed by the Clinical Data Interchange Standards Consortium (CDISC) provides a structured electronic format for organizing and exchanging nonclinical study data between sponsor companies, contract research organizations (CROs), and health authorities. In 2016, the United States Food and Drug Administration Center for Drug Evaluation and Research (US FDA CDER) started requiring the submission of SEND datasets along with nonclinical study reports for certain nonclinical study types submitted in support of commercial Investigational New Drug (IND) Applications, and New Drug Applications (NDAs). As of April 2021, over 4,000 SEND datasets have been submitted to CDER.

The CDISC-SEND data standard has created new opportunities for collaborative development of open-source software solutions to facilitate cross-study analyses of toxicology study data. FDA/CDER has established collaborative partnerships with PHUSE and BioCelerate Inc. to develop and publicize novel methods of extracting value from SEND datasets. R scripts were initially created to explore the feasibility of querying a database of SEND-formatted toxicology study datasets to retrieve historical control distributions for study endpoints (Carfagna et al., 2020). These scripts were subsequently organized and published as an R package. The package also includes an R Shiny application with a graphical user interface.

Results

Currently, the functionality of the package is primarily focused on facilitating the targeted extraction of historical control data based on user-specified study and/or animal parameters, e.g. date of study, route of administration of test article, species, animal age, etc. Additional functionality will be added to allow users to compare and contrast toxicological profiles of various test articles across studies. End users who are not familiar with the R programming language are able to utilize the R Shiny web application to perform cross-study analysis. Experienced R programmers, on the other hand, will be able to integrate the package functions into their own custom scripts/packages and potentially contribute improvements to its functionality.

Figure 2. Screenshots from the sendigR R Shiny web application. (A) Users can customize the selection of control animals by study design/date, route of administration, species, strain, and sex. (B) Line listings are generated describing the characteristics of each control animal selected. (C) Historical control distributions of toxicology study endpoints, i.e. histopathology (MI), clinical pathology (LB), and body weights (BW), are aggregated and displayed in an interactive table.

Methodology

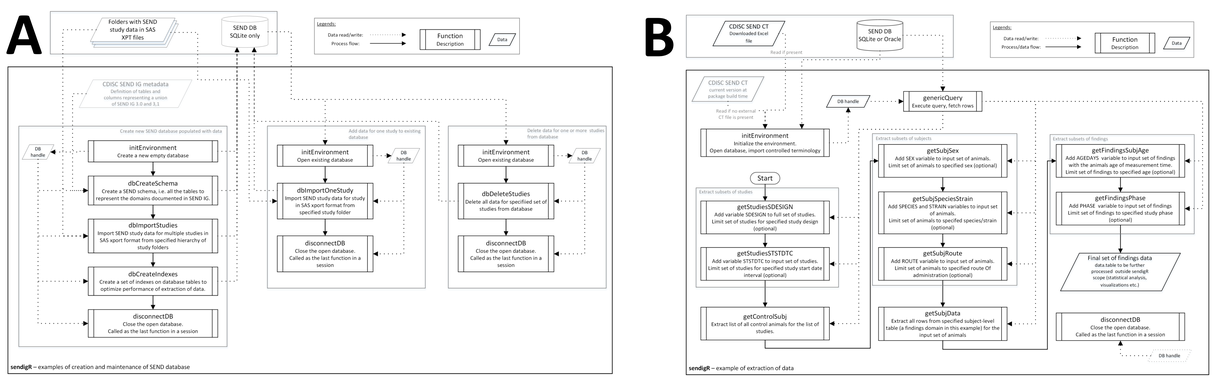

Figure 1. Schematic diagrams depicting how sendigR can be used to build, maintain, and query a SQLite database of SEND data. (A) Functions in the sendigR package can be used to generate a new SQLite database and add/remove studies from an existing SQLite or Oracle database. (B) Functions in the sendigR package can be used to query a database of SEND data to extract findings from control animals that match specific criteria.

Conclusion

The sendigR package will provide data scientists and toxicologists with a free, open source toolbox that can be utilized to interrogate large repositories of SEND-formatted toxicology study data.

Relevant Publications/Links

- A demo version of the R Shiny application is available and publicly hosted on shinyapps.io.
- The code is currently available on GitHub and the package will be submitted to the Comprehensive R Archive Network (CRAN).
- Carfagna MA, Anderson J, Eley C, Fukushima T, Horvath J, Houser W, Larsen B, Page T, Russo D, Sloan C, Snyder K, Thompson R, Ullmann G, Whittaker M (2020). Leveraging the value of CDISC SEND datasets for cross-study analysis: Incidence of microscopic findings in control animals. *Chemical Research in Toxicology*, 34(2):483-494. https://doi.org/10.1021/acs.chemrestox.0c00317