

Web Codebooks - Interactive Data Summaries for Clinical Trial Data

Jeremy Wildfire, Rho Inc., Chapel Hill, USA
Ryan Bailey, Rho Inc., Chapel Hill, USA
Spencer Childress, Rho Inc., Chapel Hill, USA
Rebecca Krouse, Rho Inc., Chapel Hill, USA

ABSTRACT

The Web Codebook tool provides a simple interactive framework for exploring data. The tool is designed to require minimal user configuration and works on any data set organized in to rows and columns. On initial load, the codebook provides a concise summary of every variable in a dataset, and then allows users to look at details using interactive features such as real-time filters, grouping and highlighting. The [source code](#), [technical documentation](#) and associated [R package](#) are available on GitHub, and a live interactive example is available [here](#).

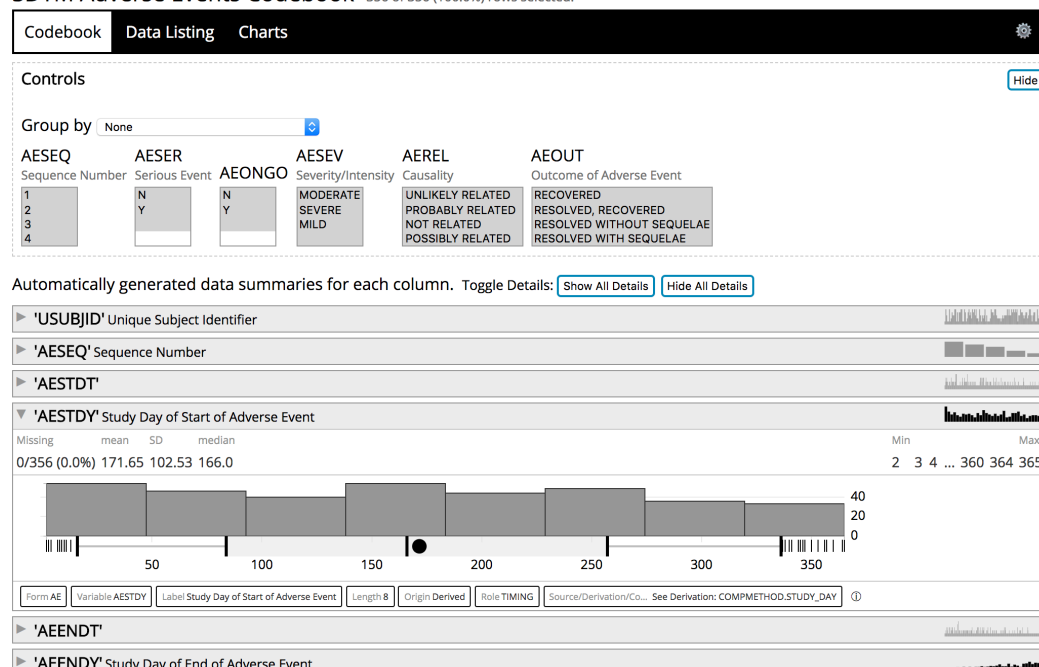
INTRODUCTION

Clinical data managers and biostatisticians are expected to have a thorough, top-to-bottom knowledge of their trial data as a means of ensuring data quality and integrity. This expert knowledge is gained by daily interaction with the data at a granular level and by comprehensive exploration of aggregate data. Data exploration in particular is a critical pre-requisite for proper interpretation of the trial results (Leek, 2015). Even though data exploration is of such importance and is performed throughout the entire life-cycle of a trial, it remains largely a technical task with a somewhat limited toolset. The web codebook is a simple, web-based interactive graphic that streamlines data exploration for any data set.

The Web Codebook tool builds upon the existing use of statistical graphics and data visualization in clinical trials by creating a simple interactive framework for exploring data. In particular, the Web Codebook takes inspiration from Frank Harrell's describe method from the hmisc R package (Harrell, 2006), and previous work in SAS (Calatroni, 2007; Rosenbalm, 2017) to produce a concise web-based data summary of every variable in a data set. Like its static codebook predecessors, the web-codebook includes paneled displays, comprehensive data listings, and charts for each variable type, but it expands on these tools by providing interactivity via dynamic filters, collapsible/expandable sections, across-chart data linking, and customizable controls. The resulting tool is well suited for use in many aspects of clinical trial research, including data exploration, anomaly detection, key end point and safety monitoring surveillance.

PhUSE US Connect 2018

SDTM Adverse Events Codebook 356 of 356 (100.0%) rows selected.



CODEBOOK DATA SUMMARIES

The web-codebook has five views that share interactive functionality. A simulated SDTM Adverse Events data set is used to demonstrate the tool's functionality below, but the tool works with any data standard and domain.

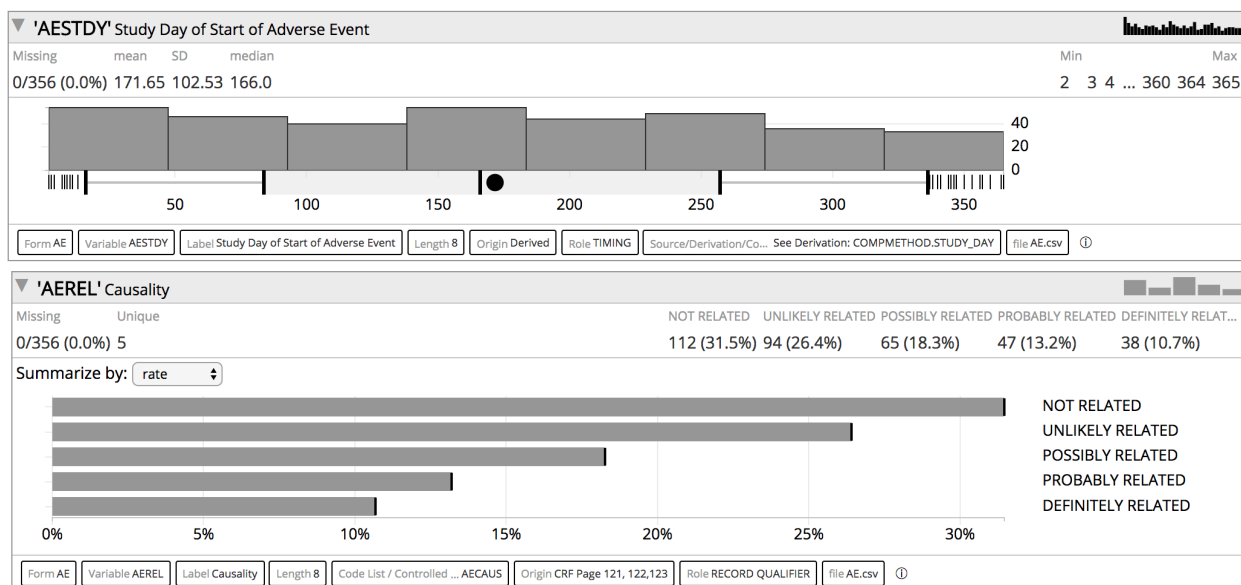
CODEBOOK VIEW

The Codebook view shows a concise summary for each variable in the loaded data set. The summaries are collapsed by default, with only the variable name, label (if any), distribution and missing data summary (if any) shown. This view provides a concise summary of the entire data set.

► 'USUBJID' Unique Subject Identifier	
► 'AESEQ' Sequence Number	
► 'AESTDT'	
► 'AESTDY' Study Day of Start of Adverse Event	
► 'AEENDT'	
► 'AEENDY' Study Day of End of Adverse Event	
► 'AETERM' Reported Term for the Adverse Event	1.4% missing
► 'AEDECOD' Dictionary-Derived Term	1.4% missing
► 'AEBODSYS' Body System or Organ Class	1.4% missing
► 'AESER' Serious Event	1.4% missing
► 'AEONGO'	1.4% missing
► 'AESEV' Severity/Intensity	1.4% missing
► 'AEREL' Causality	12.9% missing
► 'AEOUT' Outcome of Adverse Event	1.1% missing

Users can click any variable to see additional details. Appropriate summary statistics, frequency tables and charts are provided. Histograms with box plots are drawn for continuous variables (like AESTDY below) and bar charts for categorical variables (like AEREL). Variable level metadata is also shown beneath the chart if provided by the user.

PhUSE US Connect 2018



DATA LISTING VIEW

The Data Listing view provides a simple tabular output so that the user can interact with the raw data. The listing is exportable, sortable and searchable. The example below is sorted by start day.

AESTDY ↓

Search 356 records displayed

USUBJID	AESEQ	AESTDT	AESTDY	AEENDT	AEENDY	AEDECOD	AESER	AEONGO	AESEV	AEREL	AEOUT
03-013	1	2015-11-06	10	2016-06-03	220	Chronic kidney disease	N	Y	MILD	PROBABLY RELATED	RECOVERED
05-010	1	2015-12-26	10	2016-07-31	228	Azoospermia	N	N	MODERATE	NOT RELATED	RECOVERED
05-022	2	2015-07-24	101	2015-08-29	137	Cognitive disturbance	N	Y	MILD	POSSIBLY RELATED	RECOVERED
05-028	1	2015-06-09	101	2016-02-22	359	CPK increased	N	N	MILD	POSSIBLY RELATED	RESOLVED WITH SEQUELAE
05-025	2	2015-11-20	103	2016-04-19	254	Laryngitis	Y	N	MILD	PROBABLY RELATED	RECOVERED
04-009	4	2015-07-21	103	2015-11-10	215	Serum amylase increased	N	N	MILD	POSSIBLY RELATED	RECOVERED
05-029	2	2016-03-13	103	2016-09-20	294	Laryngeal obstruction	Y	Y	MILD	NOT RELATED	RESOLVED, RECOVERED
01-018	2	2016-03-25	104	2016-12-06	360	Ovarian rupture	Y	N	MILD	POSSIBLY RELATED	RECOVERED
04-004	4	2016-03-01	106	2016-10-09	328	Jejunal hemorrhage	Y	N	MODERATE	DEFINITELY RELATED	RESOLVED WITHOUT SEQUELAE
02-029	2	2016-02-04	106	2016-07-01	254	Middle ear inflammation	N	N	MILD	UNLIKELY RELATED	RECOVERED

Export: CSV

1 2 3 4 5 ... > >>

PhUSE US Connect 2018

SETTINGS VIEW

The Settings view lets users customize labels, hide variables and specify which columns should be used as interactive groups and filters, as described below.

Column	Label	Group	Filter	Hide
USUBJID	Unique Subject Identifier	☐	☐	☐
AESEQ	Sequence Number	☒	☒	☐
AESTDT		☐	☐	☐
AESTDY	Study Day of Start of Ad	☐	☐	☐
AEENDT		☐	☐	☐
AEENDY	Study Day of End of Ad	☐	☐	☐
AETERM	Reported Term for the A	☐	☐	☐
AEDECOD	Dictionary-Derived Term	☐	☐	☐
AEBODSYS	Body System or Organ C	☐	☐	☐
AESER	Serious Event	☒	☒	☐
AEONGO		☒	☒	☐
AESEV	Severity/Intensity	☒	☒	☐
AEREL	Causality	☒	☒	☐
AEOUT	Outcome of Adverse Evi	☒	☒	☐

CHARTS VIEW

The Charts view lets users interactively create simple bivariate data visualizations. The system automatically uses an appropriate visualization based on the types of the x and y variables selected.

FILE EXPLORER VIEW

The optional File Explorer view provides a simple method to load codebooks for multiple files (e.g. all analysis data sets for a study) from the same web page. Clicking a file name loads the codebook, and the user can also view the Data Listing and Charts for the selected file as desired.

INTERACTIVE FEATURES

Users can explore their data in real time using included interactive functionality using the controls shown below.

SDTM Adverse Events Codebook 356 of 356 (100.0%) rows selected. **D**

A Files

Codebook

Data Listing

Charts

Controls

B Group by

None

C

AESEQ

Sequence Number

1

2

3

4

AESER

Serious Event

N

Y

AEONGO

N

Y

AESEV

Severity/Intensity

MODERATE

SEVERE

MILD

AEREL

Causality

UNLIKELY RELATED

PROBABLY RELATED

NOT RELATED

POSSIBLY RELATED

AEOUT

Outcome of Adverse Event

RECOVERED

RESOLVED, RECOVERED

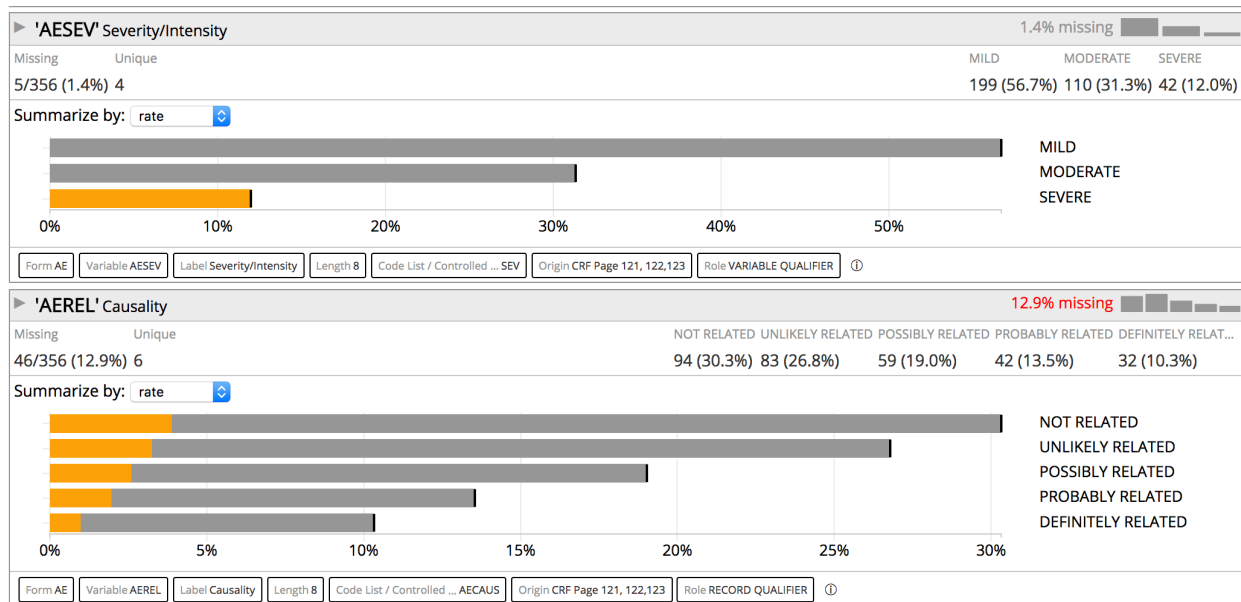
RESOLVED WITHOUT SEQUELAE

RESOLVED WITH SEQUELAE

Interactive functionality includes:

- **Navigation Bar** – The navigation bar allows users to easily move between views. (Item A above)
- **Grouping** – Any variable can be used to group the data in to strata. When a grouping variable is selected, one chart per group level is drawn for the detailed charts in the Codebook view and in the Charts section (Item B above)
- **Filters** – Filters can be created for any categorical data in the data set. Whenever a user changes the filter it is applied on the Codebook, Data Listing and Charts Views in real time. (Item C above)
- **Status Summary** – A data summary, giving number of columns, rows and highlighted records is given in the title of the codebook, and is updated whenever filters or highlighting changes (Item D above)
- **Highlight a Subpopulation** – Users can click bars in the detail view of the codebook to highlight the associated records. In particular, the participants are highlighted in other charts in the Codebook view, in the Data Listing view, and in the Charts view. The example below shows that by clicking on the “Severe” bar in the AESEV variable summary, the Severe records are highlighted in the AEREL variable summary (as well as in all other summaries, not shown).

PhUSE US Connect 2018



TECHNICAL DETAILS

The core web-codebook library is written in JavaScript using D3.js (Bostock, 2011) and webcharts.js (Bryant, 2016). In most cases, summaries for a single data set can be initialized with a single line of javascript:

```
webcodebook.createChart('#chartLocation', {}).init(data).
```

An accompanying R implementation using htmlwidgets (Vaidyanathan, 2014) and Shiny (Chang, 2015) is also available and is similarly easy to initialize, either as a standalone webpage or as a shiny application:

```
devtools::install_github('RhoInc/codebook')  
library(codebook)  
codebook(mtcars) #create webpage with htmlwidgets  
codebookApp() #run shiny app
```

The codebook is designed to work with any data file using little to no custom configuration, but many customization options for both single and multifile codebooks are provided. Full documentation is found on the [codebook configuration page](#) in the Github wiki.

Configuration of the File Explorer view is slightly more complex in JavaScript, but the R implementation provides a simple method to explore all data files in the current R session as shown in the example above using the `explorerApp()` method.

All work on the web-codebook is open source under the MIT license.

CONCLUSION

The web-codebook expands the functionality of standard static data set summaries by providing simple, web-based reports in R or a modern web browser that are open source and free to use.

REFERENCES

- Leek, J. T., & Peng, R. D. (2015). Statistics: P values are just the tip of the iceberg. *Nature News*, 520(7549), 612.
- Harrell Jr, F. E., & Dupont, M. C. (2006). The Hmisc Package. R package version, 3, 0-12.
- Calatroni A. & Mitchell H. (2007). Generating Automated Dataset Summaries, Society for Clinical Trials Annual Meeting in Montreal, QC
- Rosenbalm, S. (2018). RhoInc/sas-codebook. [online] Available at: <https://github.com/RhoInc/sas-codebook>
- Bostock, M., Ogievetsky, V., & Heer, J. (2011). D³ data-driven documents. *IEEE transactions on visualization and computer graphics*, 17(12), 2301-2309.

PhUSE US Connect 2018

Bryant, N., & Wildfire, J. (2016). Webcharts—A Web-based Charting Library for Custom Interactive Data Visualization. *Journal of Open Research Software*, 4(1).

Vaidyanathan, R., Cheng, J., Allaire, J., Xie, Y., & Russell, K. (2014). htmlwidgets: HTML Widgets for R. R package version 0.3, 2.

Chang, W., Cheng, J., Allaire, J., Xie, Y., & McPherson, J. (2015). Shiny: web application framework for R. R package version 0.11, 1(4), 106.

CONTACT INFORMATION (header 1)

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

Jeremy Wildfire
Rho
6330 Quadrangle Drive
Chapel Hill, NC 27517

Email: jeremy_wildfire@rhoworld.com
Web: <http://www.rhoworld.com/>

Brand and product names are trademarks of their respective companies.