

# Open-Source Development for Traditional Clinical Reporting

PHUSE US Connect 2021

TT08

Mike Stackhouse

Nathan Kosiba



*What if you could download an analysis reporting  
pipeline from GitHub?*

# Why Open-source

---

What did CDISC do for the industry?

Before CDISC, data structured on a company-by-company basis,  
  
now there is a common data standard

Why was this a problem to be fixed?

Companies must create  
Teams must learn  
Agency must interpret

CDISC created a unified approach

They did the work, so you don't have to

*A lot of code is redundant – why reinvent the wheel?*

# Powerful Open-source Tools



# Why can't we just adopt these tools and move our companies forward?



# pharmaRTF

Protocol: CDISCPLOT01  
Population: Intent-to-Treat

Page 1 of 3

Table 14-2.01  
Summary of Demographic and Baseline Characteristics

|                     |                 | Placebo<br>(N=86) | Xanomeline<br>Low Dose<br>(N=84) | Xanomeline<br>High Dose<br>(N=84) | Total<br>(N=254) | p-value<br>[1] |
|---------------------|-----------------|-------------------|----------------------------------|-----------------------------------|------------------|----------------|
| Age (y)             | n               | 86                | 84                               | 84                                | 254              | 0.5934         |
|                     | Mean            | 75.2              | 75.7                             | 74.4                              | 75.1             |                |
|                     | SD              | 8.59              | 8.29                             | 7.89                              | 8.25             |                |
|                     | Median          | 76.0              | 77.5                             | 76.0                              | 77.0             |                |
|                     | Min             | 52.0              | 51.0                             | 56.0                              | 51.0             |                |
|                     | Max             | 89.0              | 88.0                             | 88.0                              | 89.0             |                |
|                     |                 |                   |                                  |                                   |                  |                |
|                     | <65 yrs         | 14 ( 16%)         | 8 ( 10%)                         | 11 ( 13%)                         | 33 ( 13%)        | 0.1439         |
|                     | 65-80 yrs       | 42 ( 49%)         | 47 ( 56%)                        | 55 ( 65%)                         | 144 ( 57%)       |                |
|                     | >80 yrs         | 30 ( 35%)         | 29 ( 35%)                        | 18 ( 21%)                         | 77 ( 30%)        |                |
|                     |                 |                   |                                  |                                   |                  |                |
| Sex                 | n               | 86                | 84                               | 84                                | 254              | 0.1409         |
|                     | Male            | 33 ( 38%)         | 34 ( 40%)                        | 44 ( 52%)                         | 111 ( 44%)       |                |
|                     | Female          | 53 ( 62%)         | 50 ( 60%)                        | 40 ( 48%)                         | 143 ( 56%)       |                |
| Race (Origin)       | n               | 86                | 84                               | 84                                | 254              | 0.6477         |
|                     | Caucasian       | 75 ( 87%)         | 72 ( 86%)                        | 71 ( 85%)                         | 218 ( 86%)       |                |
|                     | African Descent | 8 ( 9%)           | 6 ( 7%)                          | 9 ( 11%)                          | 23 ( 9%)         |                |
|                     | Hispanic        | 3 ( 3%)           | 6 ( 7%)                          | 3 ( 4%)                           | 12 ( 5%)         |                |
|                     | Other           | 0                 | 0                                | 1 ( 1%)                           | 1 ( <1%)         |                |
| MMSE                | n               | 86                | 84                               | 84                                | 254              | 0.5947         |
|                     | Mean            | 18.0              | 17.9                             | 18.5                              | 18.1             |                |
|                     | SD              | 4.27              | 4.22                             | 4.16                              | 4.21             |                |
|                     | Median          | 19.5              | 18.0                             | 20.0                              | 19.0             |                |
|                     | Min             | 10.0              | 10.0                             | 10.0                              | 10.0             |                |
|                     | Max             | 23.0              | 24.0                             | 24.0                              | 24.0             |                |
| Duration of disease | n               | 86                | 84                               | 84                                | 254              | 0.1530         |
|                     | Mean            | 42.6              | 48.7                             | 40.5                              | 43.9             |                |
|                     | SD              | 30.24             | 29.58                            | 24.69                             | 28.40            |                |
|                     | Median          | 35.3              | 40.2                             | 36.0                              | 36.2             |                |
|                     | Min             | 7.2               | 7.8                              | 2.2                               | 2.2              |                |
|                     | Max             | 183.1             | 130.8                            | 135.0                             | 183.1            |                |

[1] P-values are results of ANOVA treatment group comparison for continuous variable and Pearson's chisquare test for categorical variables.  
NOTE: Duration of disease is computed as months between date of enrollment and date of onset of the first definite symptoms of Alzheimer's disease.  
Source: <run interactively>

01:11 Friday, December 04, 2020

Provide titles with  
special RTF  
formatting

Repeat column  
headers across  
pages

Control document  
attributes such as  
page size, margins,  
orientation

Footnotes with  
dates and program  
execution path



Demographic Summary: Intent-to-Treat Population

| Demographic | Parameter | Placebo<br>(N=XXX) | Active<br>(N=XXX) |
|-------------|-----------|--------------------|-------------------|
| Sex n (%)   | n         | xx                 | xx                |
|             | Female    | xx (xx.x)          | xx (xx.x)         |
|             | Male      | xx (xx.x)          | xx (xx.x)         |
|             | Missing   | xx                 | xx                |
| Age (years) | n         | xx                 | xx                |
|             | Mean      | xx.x               | xx.x              |
|             | SD        | xx.x               | xx.x              |
|             | Missing   | xx                 | xx                |

Easy  
construction  
and intuitive  
syntax

Smart string  
formatting for  
your numbers

Presentation  
flexibility  
without  
hundreds of  
parameters

| Tplyr      |              |
|------------|--------------|
| row_label2 | var1_Placebo |
| <chr>      | <chr>        |
| <65        | 14 [ 16.3%]  |
| 65-80      | 42 [ 48.8%]  |
| >80        | 30 [ 34.9%]  |
|            | 86 [100.0%]  |



# Atorus / GSK Partnership

---



Atorus and GSK aim to build, collect, and maintain a suite of R packages that support the conventional clinical reporting pipeline.



Focus of engagement is to advance open-source solutions to meet the needs and demands of the clinical statistical programming industry.



The goal of this partnership is not only to support and improve the acceptance and capabilities of R for Atorus and GSK, but for the broader pharmaceutical industry.

# Open-source Philosophy

---

|            |                                                                                            |
|------------|--------------------------------------------------------------------------------------------|
| Leverage   | If existing solutions are sufficient, leverage them where able                             |
| Contribute | If an existing solution is close to what's required, contribute to that solution (if able) |
| Create     | If no sufficient solution exists, then create one                                          |
| Build      | Build lego blocks – not a Death Star                                                       |

# Atorus / GSK Packages in Development



**timber** generates an execution log to support reproducibility and traceability of an R script



**metacore** provides an interface to read in various metadata sources and store in a standardized object



**xportr** serves as a single tool to create SAS transport files and perform pharma specific dataset level validation checks

# What Now?

---

## Communicate

- Try our packages for yourself
- Make requests for features, or report issues

## Collaborate

- Make your own updates
- Commit them back

## Contribute

- Plug a gap in the pipeline
- Create and release your own open-source packages

# Thank you

Questions?

Nathan.Kosiba@atorusresearch.com  
Mike.Stackhouse@atorusresearch.com

