Paper OS16

Teal as an Open-Source Tool Accelerates Clinical Trial Data Research at GSK

Jeffrey Long, GSK, San Francisco, CA, USA

ABSTRACT

GSK is a pioneer in embracing open-source technologies and applications. Teal, a highly engineered and validated tool developed by Roche/Genentech, is aiding the interactive exploration of data with Shiny web applications. Teal apps for RWD, machine learning output and public data provide tools for niche areas to perform exploratory investigation of data that is non-standard and outside the scope of our data visualization tool RAPIDO where the focus is CDISC and Analysis Results Data. Teal is enabling the refinement of patient selection in oncology trials, and supporting the generation of new hypotheses as biomarker and baseline data are interrogated in new ways. GSK's move to a cloud-based platform has enabled custom Docker runtime environments that expedite the installation of {teal} packages and dependencies. These pre-built environments enable end-users to rapidly deploy the functionality-accelerating exploratory, early development, safety, and efficacy analyses, furthering the goal of improved outcomes for patients!

INTRODUCTION

GSK leverages open-source tools to accelerate clinical trial insights. As a prime example, Teal Apps from R-based packages have been developed by the Insights Engineering organization to explore interactive data analyses. Currently projects in Oncology Disease Area Strategy Statistics, Oncology Biomarker Programming/ Translational Statistics, SPC Biomarker Programming/ CPEM Statistics, and Vaccine Research Translational Statistics teams are using Teal Apps. The data agnostic framework and decoupling data from analysis modules has provided new ways of looking at exploratory, longitudinal, and clinical datasets. Teal Apps can be built for SDTM and ADaM datasets, machine learning output from RWD, cell-type-specific RNA expression, protein quantification and preclinical research data. The collaborative data sharing efforts are using Teal Apps to glean insights from TransCelerate EMD Serono, Eli Lily, Roche, Flatiron, vaccine development pipeline raw data and more.

Teal Data Visualization Use Cases at GSK

A Patient Profiling Teal App displays safety monitoring visualization for longitudinal vital signs, therapies, adverse events, and lab test data. Choosing parameters such as blood pressure, pulse rate, temperature, respiratory rate, weight, and height gives change in readings across visits. Similarly, lab tests over time can be plotted for each patient (e.g., leukocyte count, erythrocytes, hematocrit, platelet, bilirubin, creatinine, potassium, albumin, calcium, glucose, etc.). Then adverse events can be visualized on a timeline (e.g., epidermal, and dermal conditions, gastrointestinal motility conditions, hepatobiliary investigations, appetite and nutritional disorders, vascular hypertensive disorders, or respiratory disorders, infections, etc.).

Six Teal Apps were published to GSK Posit Connect servers for all researchers to access and explore random CDISC data and capabilities. These included:

- Patient Profile Teal App for Safety Monitoring
- Longitudinal Clinical Trial Data Analysis Teal App
- Early-Phase Clinical Data Analysis Teal App

- Exploratory Teal App
- Clinical Trial Safety Teal App
- Clinical Trial Efficacy Teal App

Researchers, statisticians, and clinicians can explore the diverse data analysis modules available for application to their data.

Clinical Trial Efficacy Teal App

The Clinical Trial Efficacy Teal App allows one to explore survival data by changing stratification variables, biomarker selection, and patient cohorts. Choose from data table interactions, variable browsing demographic tables, forest plots, Kaplan Meier plots, response plots, time to event tables, cross tables, regression analysis, MMRM, binary response or ANCOVA in a single Shiny application. The ability to facet KM plots on a categorical biomarker or a derived discrete variable from a continuous biomarker within the application delivers extreme flexibility.

Translational Statistics Longitudinal Laboratory Analysis Teal App

Translational Statistics Longitudinal Laboratory Analysis Teal App was developed for most of the vaccine research studies which are immunogenicity studies. This Teal App aids in the search for the formulation that maximizes the immune response in a multidimensional way: antibodies and cell mediated immunity while not producing too much reactogenicity. Endpoints and response are measured at multiple time points. These readouts are then analyzed in the Teal framework by vaccine formulation, material, mRNA, and immunogen. Data engineers design connectors for our data to fit into generic Teal Apps for rapid deployment. Visualizations of Elisa protein results by timepoint in box plots, line plots and spaghetti plots can be facetted by vaccine metadata.

Flatiron Immunotherapy Patient sidClustering Output Teal App

At GSK, our statisticians have applied the published sidClustering algorithm[1] to Flatiron stage 4 lung cancer patients with prior immunotherapy. This unsupervised random forest machine learning algorithm grouped patients into 4 clusters. The associations with each cluster provide insights on patient selection criteria. The association of ECOG (Eastern Cooperative Oncology Group) values, histology and treatments with cluster identification can be visualized in our Teal App.

[1] Reference: Mantero, A., & Ishwaran, H. (2021). Unsupervised random forests.

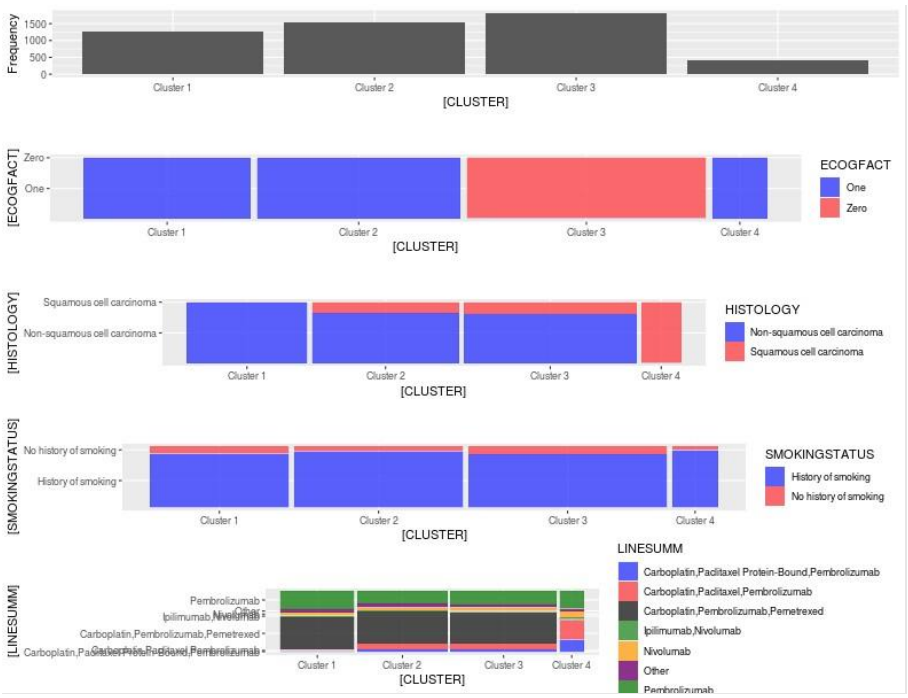

Figure 1: Flatiron Teal App association plot showing ECOG value, histology, smoking status, and proportion of line treatment for patients clustered by sidClustering algorithm.

These associations (Fig. 1) inform our selection for facets in a survival analysis using the R package {teal.modules.clinical} function `tm_g_km()`, as seen below in (Fig 3,4). An R code snippet for the Kaplan Meier plot (Fig.2) shows the organization of parameters for the teal module function. The strategy of decoupling data from Teal App functions allows reproducible, concise, human-readable code blocks.

main teal02 / app.R

Code Blame Raw

```
487       tm_g_km(
488         label = "Kaplan Meier Plot",
489         dataname = "ADTTE",
490         parentname = "ADSL",
491         arm_var = cs_arm_var,
492         paramcd = cs_paramcd_tte,
493         strata_var = cs_strata_var,
494         facet_var = cs_facet_var,
495         time_unit_var = teal.transform::choices_selected(teal.transform::variable_choices(ADTTE, "AVALU"),
496                                                          "AVALU", fixed = TRUE),
497         aval_var = teal.transform::choices_selected(teal.transform::variable_choices(ADTTE,
498                                                                                      "AVAL"), "AVAL", fixed = TRUE),
499         cnsr_var = teal.transform::choices_selected(teal.transform::variable_choices(ADTTE,
500                                                                                      "CNSR"), "CNSR", fixed = TRUE)
501       ),
```

Figure 2: Flatiron Teal App module code for Kaplan Meier plot function `{r} tm_g_km()`.

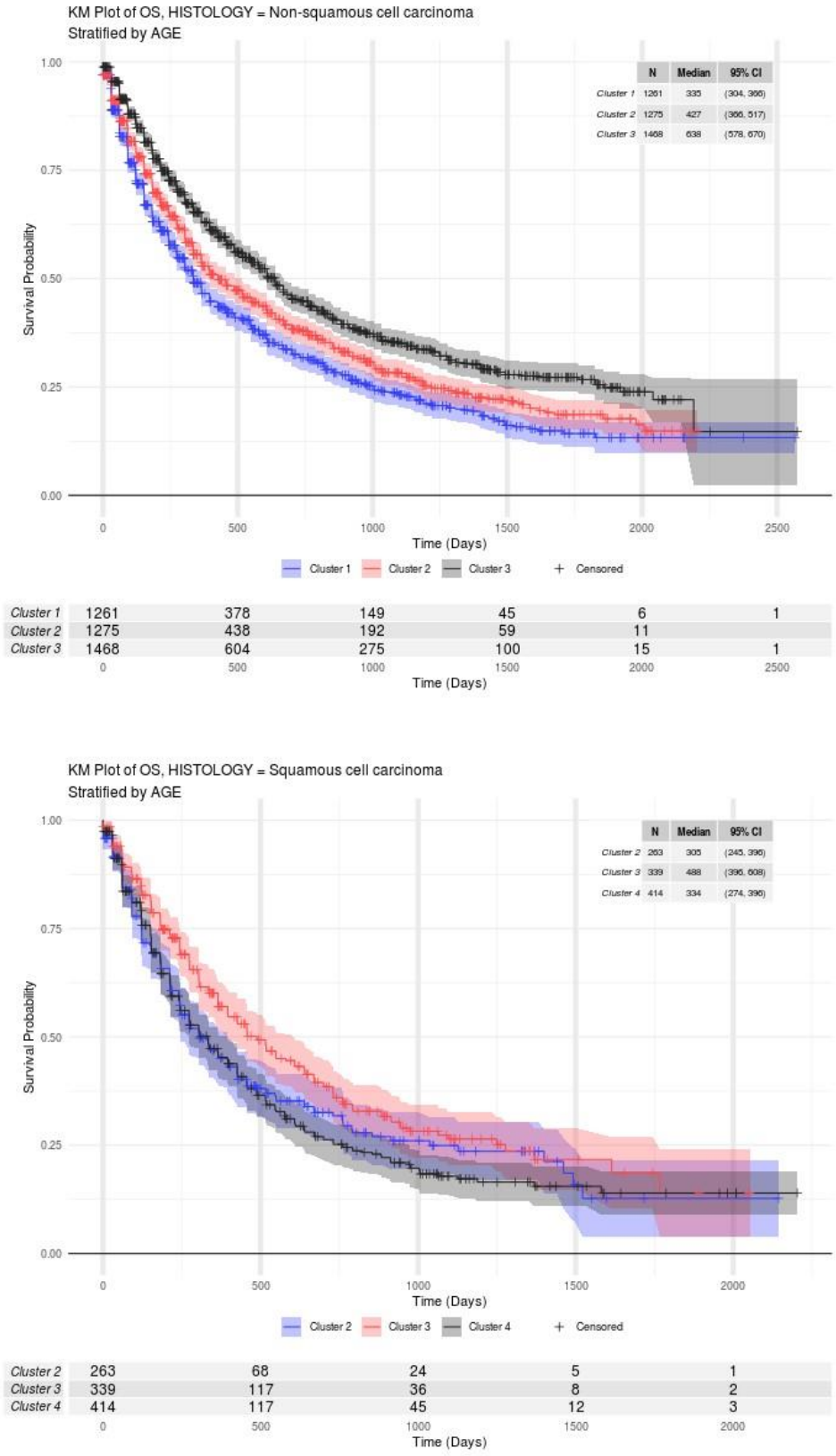

Figure 3: Kaplan Meier Plot of overall survival from Flatiron sidClustering Teal App faceted on lung cancer patient histology, top for non-squamous cell carcinoma and bottom for squamous cell carcinoma.

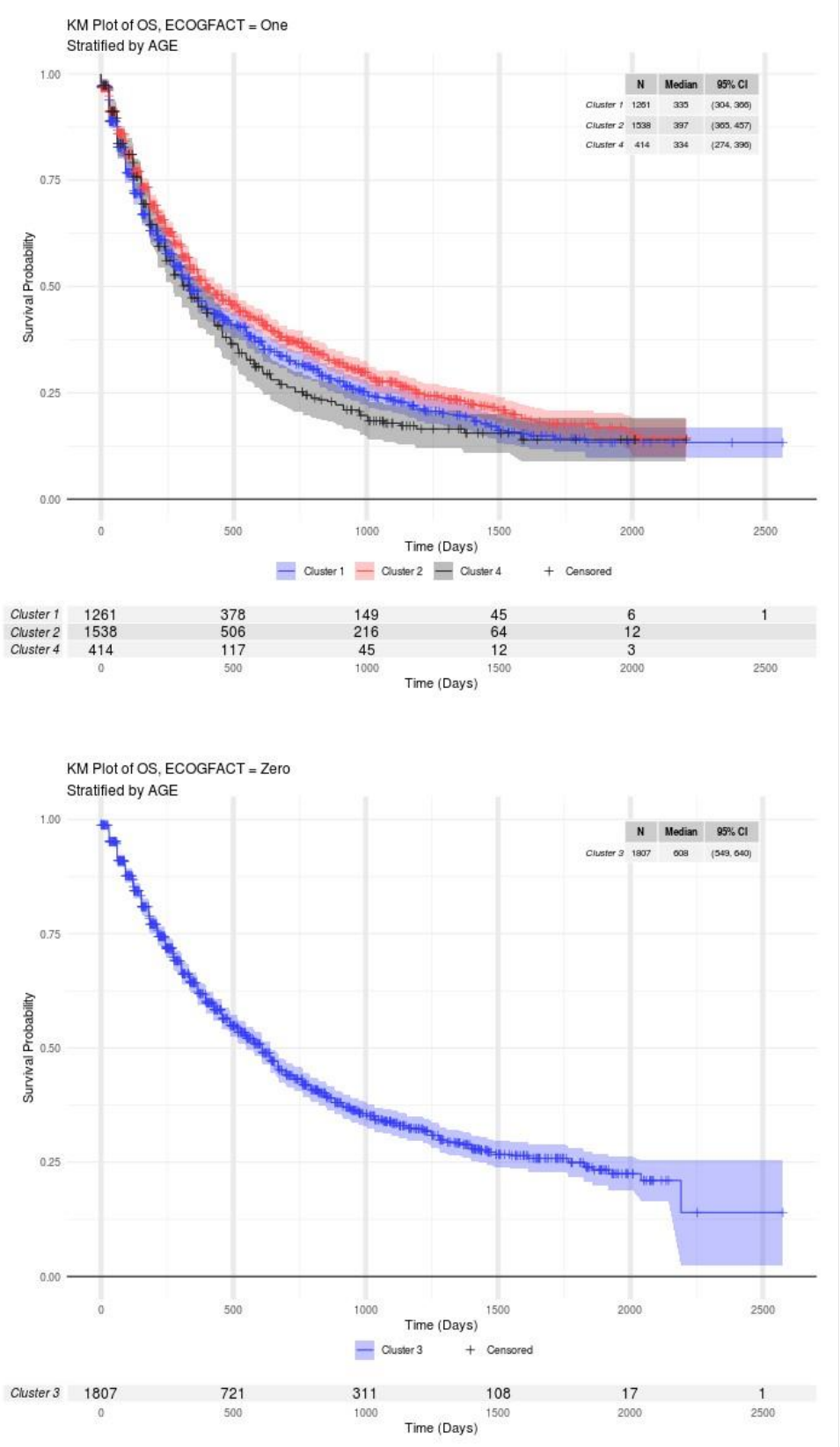

Figure 4: Kaplan Meier plot of overall survival from Flatiron sidClustering Teal App with figure faceted on ECOG value, with ECOG = 1 on top and ECOG = 0 on bottom.

Further Insights Engineering Open-Source Resources Used at GSK

In addition to {teal} package development, the Insights Engineering[2] organization has open sourced a Quarto document, the Biomarker Analysis Catalog[3]. If you are looking to perform a static analysis of biomarkers with Kaplan-Meier, Response or Survival Forest graphs, this resource provides R code templates for everything you need to get started. It walks through the code to create random CDISC datasets and the plots for each analysis. Along a similar line, the TLG Catalog[4] is an effort to deliver R code templates for tables, listings, and graphs common to clinical reporting. The TLG Catalog also aims to deliver Teal App code for figures in the Graphs category. However, this aspect of the project is still in progress.

Deployed to SPACE

GSK has developed and transitioned to a cloud-based Statistics and Programming Analytics Computing Environment (SPACE). Our canonical data visualization tool, RAPIDO has been tuned for clinical trial output and is adopting many facets of the Teal framework. The GSK Oncology Data Strategy team has developed a simple loading design for {teal} packages via Domino and Docker runtime environments in our corporate cloud resource called SPACE. This enables use of large datasets, umbrella studies, and biomarker study-aligned data analysis with no foreseeable constraints. This strategy has enabled robust data visualization for clinical and preclinical research data re-use.

One last resource to mention is our GSK R Snippet Quarto Blog. Our statisticians are excited to have a new crowd sourced internal publication process to share code, analysis vignettes, and helpful news. The self-service deployment has allowed our employees to post updates to the teams via an enterprise GitHub repository.

CONCLUSION

As open-source tools in pharma are shared by statistical sciences, we are encouraged to stand on the shoulders of giants as we move ahead together. This is supercharging our data to improve outcomes for patients. Thank you for listening and thanks to the great teams at GSK and Insights Engineering.

REFERENCES

1: Mantero, A., & Ishwaran, H. (2021). Unsupervised random forests. *Statistical analysis and data mining*, *14*(2), 144–167. https://doi.org/10.1002/sam.11498
2: Insights Engineering https://github.com/insightsengineering
3: Biomarker Analysis Catalog https://insightsengineering.github.io/biomarker-catalog/
4: TLG Catalog https://insightsengineering.github.io/tlg-catalog/stable/

ACKNOWLEDGEMENTS

Rebecca Greenwood, Tripti Bhaskaran, Timothy Fantuzzo, Ilana Trumble, Antara Majumdar

[2] Insights Engineering https://github.com/insightsengineering
[3] Biomarker Analysis Catalog https://insightsengineering.github.io/biomarker-catalog/
[4] TLG Catalog https://insightsengineering.github.io/tlg-catalog/stable/

RECOMMENDED READING

A Gallery of Exploratory Web Apps used for Analyzing Clinical Trial Data, https://github.com/insightsengineering/teal.gallery

Sabanes Bove D, Moroni P, Rossomanno S, Luong J, Bhatia N, Rossomanno S, Li H, Midori Zhang L, de la Rua E, Banjo S, Falquet B (2023). biomarker.catalog: Biomarker Analysis Catalog. R package version 0.2.0, https://insightsengineering.github.io/biomarker-catalog/.

CONTACT INFORMATION

Jeffrey Long
GSK
62 Buena Vista Terrace, San Francisco, CA 94117
+1 (415) 515-8082
jeffrey.x.long@gsk.com
Contact us | GSK