

R-Burst: The User Support Approach Shaping GSK's Open-Source Strategy

Andy Nicholls, GSK, London, UK
Shannon Haughton, GSK, London, UK

ABSTRACT

GSK Biostatistics has an objective to ensure 50% of our code is written in R by the end of 2025. After six years of developing open-source training and tooling, the Data Science Innovation and Engineering team shifted our strategy to focus entirely on our end users through the "R-Burst" initiative.

The initial three-month burst of user hyper-care addressed immediate challenges in adopting open-source today. We supported Programmers, Statisticians, and Data Scientists by saying "yes" to all support requests. Our only requirement was that we work together on solutions, and share the resulting insights through our daily blog for the benefit of the entire Biostatistics community. This user-focused strategy has helped us overcome adoption hurdles and informed our open-source innovation beyond 2025.

INTRODUCTION

In 2023, GSK's Biostatistics leadership decided to commit to an open-source future. Specifically, they made two major commitments:

- All central tools to be built using open-source languages
- 50% of GSK Biostatistics' code to be written in open-source languages by end 2025

With clear direction from senior leadership, the Data Science Innovation and Engineering (DSI&E) Team set a new objective to complete the open-source clinical reporting pipeline by filling in the remaining tooling gaps by the end of 2024. With the mission completed by mid-2024, attention turned towards mass user adoption. The result was a new initiative called R-Burst which began in early 2025.

A BRIEF HISTORY OF OPEN-SOURCE IN GSK BIOSTATISTICS

GSK Biostatistics started looking at open-source options for clinical reporting in 2017. The journey began with the creation of new internal training materials focused on Tidyverse with a graphics extension course dedicated to {ggplot2}. In parallel, work began on "WARP", GSK's first dedicated R platform, built around the RStudio IDE with the ability to publish *Shiny* apps and *Rmarkdown* via RStudio *Connect* (now Posit *Connect*). Following significant contributions to both the R Consortium (Nicholls, Bargo, Sims, & R Validaiton Hub, 2021) and TransCelerate (TransCelerate, 2023) white papers on risk-based assessments, GSK Biostatistics released an internal "R4GxP" processes for assessing R package quality. This paved the way for the use of R in clinical reporting at GSK.

Fast forward to 2023. GSK's R adoption journey is in full flight with several of GSK Biostatistics' R packages now publicly available via the [Pharmaverse](#). Within GSK, R is routinely being used outside clinical reporting in research settings and for trial design simulation. Within the clinical reporting space, a few ambitious study teams have started using it for production, although it is still mostly limited to graphics and for QC work. Despite limited use, most of the key questions around, "can we use R effectively?" have been answered. Following an in-depth discussion about GSK's future plans with R, the Biostatistics leadership decided to commit to an open-source future. Specifically, they made two major commitments:

- All central tools to be built using open-source languages
- 50% of GSK Biostatistics' code to be written in open-source languages by end 2025

THE DATA SCIENCE INNOVATION AND ENGINEERING TEAM

The Data Science Innovation and Engineering (DSI&E) team was formed in 2018 with the goal of building Data Science capability across the wider Biostatistics community. The team has grown to include around 15 permanent staff, plus a handful of contractors. The DSI&E team's current remit includes innovation in systems, tools, training and user support. Working in conjunction with Clinical Programming's Technical Excellence team and motivated individuals from the pipeline, the DSI&E team has been behind many of the open-source key milestones at GSK, including the creation of GSK's internal training materials; a first analytics platform for R platform; the creation of the *R4GxP* process; the vast majority of GSK's packages on Pharmaverse; and [AccelerateR](#), whereby a dedicated agile squad works closely with new study teams to accelerate their adoption of R.

INTRODUCING "R BURST"

With Biostatistics having committed to an R-based future towards the end of 2023, the DSI&E team set an internal objective to fill in the remaining R-tooling gaps for clinical reporting by the end of 2024. By mid-2024, the DSI&E team had met this objective, releasing the [{docorator}](#) package, which adds the finishing touches to displays (adding titles, footnotes, etc. and generating a corresponding PDF or RTF).

Following the release of {docator} the team spent the remainder of 2024 working with early adopters to address any teething problems. But, in anticipation of a new stretch objective to achieve 70% of GSK Biostatistics' code written in open-source languages by end 2025, DSI&E leadership began discussing a new approach to support mass user adoption. The new idea, based on a short term "burst" of energy to support the adoption of R, was quickly labelled "R-Burst".

INNOVATION PHILOSOPHY: WORKING CLOSELY WITH THE PIPELINE

It's almost impossible for a developer to be successful if they do not understand the workflows of the people using their tools. The DSI&E leadership believes that working closely with the pipeline is vital in shaping future innovations. In addition to bringing in users at an early stage of development, the DSI&E leadership is always actively seeking opportunities to promote the cross-fertilisation of ideas that comes from working closely with the pipeline. As an example, junior DSI&E staff are encouraged to take short-term secondments into the pipeline. And in a previous R adoption initiative, AccelerateR, DSI&E team members joined the study teams for four, 2-week sprints to help them use new tooling.

LET'S JUST SAY, "YES"!

R Burst was born when we asked the question, "what if we just said yes to everything for a while?" Since its creation the DSI&E team had been inundated with requests for support from various teams who wanted to tap into the team's depth of R knowledge. These requests varied from complex questions about traceability and logging to every-day R coding questions such as, "how can I adjust the positioning of my confidence intervals?" and "why is my simulation running slowly?" Despite a strong desire to help everyone, the core remit of the team is to provide innovative solutions and basic R questions were passed on to other teams. More bluntly, the team had become accustomed to saying "no"!

Moreover, a team that says "yes" to everything can be quickly overwhelmed. But saying "yes" to support requests for a short while might be possible, provided the team stopped everything else. The DSI&E leadership team therefore agreed that from early 2025 (almost) all package development work would be put on hold. Everyone available would spend three months saying "yes" to incoming support requests. Further, the leadership team would actively reach out to all users to encourage them to submit and support request that they felt appropriate.

A DAILY BLOG

Saying "yes" to support requests was the first of two core components of R-Burst. The second was a daily blog. The idea of the blog format was to create a lot of supporting information, quickly. Formal documentation tends to get bogged down in (often necessary) review process. The blog would enable the team to provide information about a wide range of topics at varying levels of complexity. Short "FYI" posts could remind people of existing tools, processes and documentation. The team also aimed to summarise any interesting support requests via the blog. This would help demonstrate impact whilst, hopefully, reducing the potential for repeat requests around the same topic. Further posts that introduced new, or more advanced topics would form the basis of more formal guidance. To that end, blog post engagement would be tracked to help prioritise topics of most interest/relevance to the user base. As a secondary benefit, the blog was a way of making some noise. As part of a change management strategy, the DSI&E team committed to daily posts. These daily posts would be summarised weekly and posted via GSK's internal social media. This would help remind everyone of the initiative.

RUNNING R-BURST

From a user's perspective, the initiative was centred around a very simple hub/landing page, built using [Quarto](#). This included the blog, a couple of sentences summarising the initiative and information about how to get in touch. Getting in touch was made as simple as possible for the user. No forms were required, they simply needed to send an email to a dedicated mailbox that reached the entire R-Burst team.

In the backend, a GitHub Project was used for project management. This same project was used to manage both the incoming support requests and the website/daily blog. The DSI&E team had already built tools that apply a GSK theme / colour scheme to a variety of Quarto-based output and the website was up and running within a couple of days.

MAXIMISING IMPACT

The R-Burst initiative was planned to start during Q1 2025 to align to the release of the new stretch objective of 70% code written in open-source languages by the end of 2025. Thus, Biostatistics leadership were able to announce the target and, at the same time, provide teams with a concrete support model to help meet the objective.

Whilst the team built the website and configured the GitHub project, the R-Burst leadership set up time with each of the various leadership teams across GSK Biostatistics in order to promote the upcoming initiative. Further promotion was conducted via GSK's internal social media, as well as via Biostatistics leadership communication (email, townhalls, etc.) The messaging across all these touch points was very consistent – 'you might be worried about your new stretch objective, but we're here to help in any way we can!' Emphasis was placed on 'any' – it was made clear that nothing was out of scope and the website was amended to include an example of set of questions.

MANAGING REQUESTS

When a new request came into the mailbox it was triaged by the R-Burst team. During the project setup, team members were asked to centrally list their specific areas of expertise. The aim was to respond to every request within

24 hours and if a request was not voluntarily picked up the list was used to appropriately triage to individuals with the relevant expertise by the R-Burst lead.

For straightforward issues, a simple email solution was sent. For more complex questions, an initial reply informed the requestor, and the rest of the R-Burst team, that further discussion would be required via *Teams* chat or through a scheduled call. The assigned team member then maintained regular communication with the requestor until the issue was fully resolved. For tracking, each request that came in was logged as an issue in our GitHub repository. This action was performed by the R-Burst team to minimise the effort required by the user. This allowed the R-Burst team to collect essential information, including the requestor's name and email, the area of Biostatistics represented, the associated study (if applicable), and an indication of the effort required to address the issue. These records not only allowed the team to follow each request through to completion, but to later identify common themes. This, in turn, informed updates to FAQ and TFL documentation and helped prioritise the development of additional resources to bridge the gaps as they appeared. The information was also used to conduct follow-up surveys with all requestors, to gather feedback on how the support model could be improved.

CASE STUDY: COX PROPORTIONAL HAZARDS IN R

A study team was performing a time-to-event analysis using Cox proportional hazards with the exact method for handling ties. Their production code was written in SAS, and they wanted to produce their QC code using R. Although they had been able to match results between SAS and R using other methods for ties, it hadn't been possible to match the required 'exact' method and so they reached out to the R-Burst team. The request was triaged across the department to those with the most expertise in this area, and after a detailed review of the two languages, the issue was identified. It was determined that SAS's 'exact' method employs the exact marginal likelihood, whereas R's 'exact' method utilises the exact partial likelihood (corresponding to the "discrete" method in SAS). Using this information the study team were able to use a consistent method for both production and QC and get a match on their results.

It was felt that the problem was likely to be encountered by other users looking to use a combination of R and SAS within the industry, so the team fed the example back to [CAMIS](#), a collaborative industry effort to capture statistical differences between software. Thus, the team had provided in-time support to the team, and ensured a lasting solution was available to others both at GSK and more broadly across industry. A brief blog post was written for the example and the engagement with CAMIS was highlighted to leadership teams as a desired behaviour when working in R/open-source.

MAINTAINING THE BLOG

Work started on a blog weeks before R-Burst officially kicked off. The R-Burst team were aware from prior experience that it would become difficult to maintain a regular blog once the support requests started flooding in. Over a period of 2-3 weeks, the team managed to build up a backlog of around 20 posts that would be released to a specific schedule once R-Burst had officially begun.

The blog had its own GitHub repository, but the review process was managed via the same GitHub project as the support requests (with labels used to identify, and subset issues where required). To create a new post, a user would branch the repo, create their post and submit a GitHub Pull Request (PR). The PR process was set up such that each PR was assigned a check list of actions for both the author and reviewer. For the author, this included basic checks to ensure the post was clean (e.g. just the '.qmd' file and any images, not any HTML output) and followed appropriate naming conventions. For the reviewer, expectations had already been set that it was not a 'wordsmith' review. Rather, the review was focused on alignment to expected ways of working, and cross-referencing other resources. For example, are packages explicitly referenced in code chunks using `'library'`, and has the author stated whether or not they are available (soon to be available) for use within GSK's GxP workflows? Are links to further information provided?

Once the PR review was complete, the merging was overseen by a small team of three individuals. They accepted the PR's to a daily schedule. An automated Posit *Connect* pipeline would then pick up the changes and publish. The same team that oversaw the PR's was responsible for additional communication to the wider Biostatistics community in the form of a weekly round-up. This was released via internal social media each Wednesday to maximise engagement.

HOW DID IT GO?

The primary aim of R-Burst was to make users feel supported. To that end, the initiative was a major success with 58 separate requests supported with 100% satisfaction reported in the user survey. The secondary objective of publishing a blog post per day was also met.

USER SUPPORT

Over three months, the team supported 58 individual requests, approximately one per working day. Around half (27) of these came from the Programming group, with the rest spread amongst the Clinical and Non-clinical Statistics groups, and the Real-world Data team that make up the GSK Biostatistics group.

Other than the volume of support, the primary success measure was user satisfaction. Working in an agile manner, the team sought feedback throughout the initiative. This was mostly verbal, or ad hoc via email. But for the purposes of demonstrating impact to senior leaders, the team also sent a survey to everyone they had supported during the first three months. Expecting a low response rate, the survey was kept short (to one page) and reminders were sent

highlighting the exact number of questions to expect (as opposed to, “this will take you no more than five minutes to complete”). The response rate of 34% was considered a success.

Satisfaction was measured on a standard, five-point Likert scale ranging from “very dissatisfied” to “very satisfied”. Of 22 responders, 21 responded to say that they were “very satisfied” with the one other response falling in the “somewhat satisfied”. Even with a relatively high response rate, such surveys should always be treated with caution. Typically, responses come mainly from those at either tail of the satisfaction distribution. People are either very happy, or very unhappy! And those in the middle are often underrepresented. Even so, the lack of dissatisfied users in such an impactful change initiative was very encouraging.

BLOG

Alongside the user support, the team managed to meet its secondary objective of producing a blog every working day for the three-month experiment. In all, 64 posts were written and published on the specially created quarto website. The blog page received >800 unique views (16,000 total views) with ~90% returning. User data were tracked using Posit *Connect* and with the `get_usage_static` function from the `{connectapi}` package used to generate an initial table of information.

SHAPING FUTURE STRATEGY

Whilst the core aim of R-Burst was to support users, the DSI&E leadership were keen to use it to learn more about how the team’s tools were being used by the pipeline, and to identify new challenges that might require future innovation. Again, the initiative can also be considered a success. The majority of the support requests were straightforward to resolve, but one or two highlighted small gaps in expected functionality, where the toolset did not fully align to either new or existing standards. In addition, the `{docorator}` case study below highlighted a need for a major change to the tooling to support a workflow that was otherwise completely unsupported.

CASE STUDY: THE `{DOCORATOR}` PACKAGE

The DSI&E team had been developing an R package called `{docorator}`, which was originally designed around a single primary function, `docorate`. This function was intended to process outputs generated using the `{gt}` and `{tfrmt}` packages and render them into PDF documents. However, a problem was identified when the R-Burst team received a request regarding an additional output format, `rtf`, which is still required for GSK’s Medical Writing function downstream.

In response, the DSI&E team refactored the package by splitting the functionality of `docorate` into two distinct components:

- `as_docorator`. This function takes the table object and any document decoration such as headers and footers, and prepares it for rendering.
- `render_pdf` and `render_rtf`. These functions can be piped together with `as_docorator` to produce our multiple output types.

This modular approach allowed the team to implement a dedicated function to address immediate needs, leaving the package open to support additional output formats in the future with more `render_*` functions.

ADAPTING R-BURST

About a month prior to the official end of the R-Burst experiment, it was decided that R-Burst should continue. However, the approach needed to be adapted to reduce the burden on the DSI&E team and enable developers to return to developing tools.

The first major change was to the blog post frequency. Overall, the blog had worked well, but it was not completely without issue. Users were asking for information on more complex topics, and it was proving difficult to produce the relevant blog posts of sufficient quality within the timeframe. On a handful of occasions, posts had gone out with minor issues which were later retrospectively corrected. And in one case, a post was almost entirely re-written post hoc. The decision was therefore made to roll back the blog with a current rate of approximately one post every 7-10 days.

The support approach was adapted to roll back some of the more time-consuming elements of the user-centric approach. In the first three months, the team made a point of starting a *MS Teams* conversation with most users to make sure they felt fully supported on an individual level. For the more complex requests, this approach continues, but for simpler requests the team keeps the discussion within a brief email thread, pointing towards relevant documentation as much as possible. To that end, R-Burst support requests have been used to prioritise the documentation that teams need. The Template Hub is one such example.

CASE STUDY: TEMPLATE HUB

Transitioning to an R-based workflow presents significant challenges for those with years of experience using SAS. It is often unclear where to begin when confronted with a new R session, multiple packages, and the requirement to produce standard output, similar to those produced by SAS-based tooling. The R-Burst team had noticed a trend in questions about standard displays, and decided to create a small library of reference examples. The library evolved organically with a focus on displays that were most commonly required. FAQs and additional examples were added as new requests came in. The library has proved to be an incredibly valuable resource to Programming teams, with >5,500 views in September alone.

Strong user engagement with Template Hub examples has enabled the DSI&E team to identify gaps in the tooling. During the initial R-Burst experiment, the team opened multiple issues in open source packages such as `{gt}` and

{tfrmt}, which Biostatistics uses for table creation. This provided a clear list of priorities for the team to work on, once the three months were over and the team could return to package development. Notably, 50% of the bug fixes and features in the latest release of {gt} (1.1.0) were authored by the GSK team.

CONCLUSION

An open-source, user-centric culture is not just a 'nice to have' for engineering teams. The successful adoption of tools *depends* on proactive user-engagement. R-Burst was a successful experiment which helped strengthen the bond between the DSI&E team and the wider Biostatistics department. The impact of saying “yes” to users benefitted both the users themselves, and the Data Science Innovation and Engineering team. The users felt fully supported by the approach, enabling them to adopt R without fear of major blockers delaying their core deliveries. Meanwhile, R-Burst helped the DSI&E team identify and prioritise various user challenges. Further, the discussions sparked new innovation ideas that the team can take forward in 2026.

While the short-term benefits of R-Burst were clear, broader success needs to be assessed over a long-term. The continuation of the initiative beyond the initial three months was an acknowledgement of the desire from users to have access to subject matter experts. Moving forward, the DSI&E team will need to strike the right balance between saying “yes” and the team’s core responsibility to provide innovative technical solutions.

REFERENCES

Nicholls, A., Bargo, P. R., Sims, J., & R Validaiton Hub. (2021, January). *A Risk-based Approach for Assessing R package Accuracy within a Validated Infrastructure*. Retrieved from R Validation Hub: <https://pharmar.org/white-paper/>

TransCelerate. (2023, October). *Modernization of Statistical Analytics Solutions*. Retrieved from <https://www.transceleratebiopharmainc.com/assets/modernization-of-statistical-analytics-solutions/>

ACKNOWLEDGEMENTS

Thanks to all of those within the DSI&E team who helped to ensure the success of R-Burst. Thanks, in particular, to Becca Krouse and members of the core R-Burst team who have been dealing with the support requests. Thank you finally to the users who have given R a try and reached out to the R-Burst team when necessary.

RECOMMENDED VIEWING

For further background to GSK Biostatistics’ journey with R, it is recommended to watch the [Posit Webinar, "GSK's R Journey: From Pilot Projects to Enterprise Adoption"](#).

CONTACT INFORMATION

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

Author Name: Andy Nicholls

Company: GSK

Address: GSK HQ, 79 New Oxford Street, London WC1A 1DG

Email: andy.p.nicholls@gsk.com

Author Name: Shannon Haughton

Company: GSK

Address: GSK HQ, 79 New Oxford Street, London WC1A 1DG

Email: shannon.l.haughton@gsk.com

Brand and product names are trademarks of their respective companies.