Paper SA01

E2E R Submission from Roche Oncology Study

Jingyuan Chen, Genentech, South San Francisco, USA
Hinal Patel, Roche, Welwyn Garden City, UK

ABSTRACT

In the rapidly evolving landscape of pharmaceutical development, Roche embraces the use of open-source R in the regulatory submission process. In this paper we will share a pioneer experience of end-to-end R submission for a new drug application to the FDA, EMA and the NMPA. We will unfold the journey of overcoming the initial challenges to achieving the filing goal and emphasise the practicalities of preparing clinical trial data in R leading to a comprehensive eSubmission package. We will also share our strategic communication and collaboration with regulatory agencies, highlighting how R-generated data and documentation were effectively presented and validated within the submission framework. The successful end-to-end NDA submission using R packages, the majority being open-source packages (from Pharmaverse), not only demonstrates the feasibility and benefits of leveraging such a powerful tool in the regulatory realm but also sets a precedent for its wider adoption in the pharmaceutical industry.

INTRODUCTION

In recent years, open-source collaborations have significantly advanced drug development within the pharmaceutical industry. Companies are increasingly adopting R, investing in upskilling their teams with new tools and systems. Cross industry initiatives such as Pharmaverse have established an R backbone for clinical trial reporting, while working groups within the R consortium have facilitated many collaborative submission pilots with the FDA (Food and Drug Administration). The outcomes of these projects have provided the foundations and solutions to help organisations prepare for a new chapter in regulatory clinical trial reporting. This paper details our journey as early adopters - the first at Roche, to leverage open-source R packages from Pharmaverse specifically for the preparation of SDTM, ADaM datasets, and the creation of tables, listings and graphs (TLGs) for regulatory submissions to health authorities. Piloting Roche's new computing platform OCEAN (One CEntralised ANalytics platform): a state-of-the-art, language agnostic, cloud-based Analytics & Data platform for both regulatory and exploratory clinical trial work brought to us by NextGen Tools & System team within Roche PD Data Science.

Our end-to-end approach using the Pharmaverse suite of R packages has allowed us to streamline and modernise our submission process. The aim of this paper is to share insights into our programming approach, the challenges and successes we encountered during the R submission process, and the interactions and dialogues with regulatory agencies. Additionally, we will highlight the supplementary content and considerations to ensure a smooth and successful R submission.

EARLY ADOPTERS JOURNEY

As early adopters we assembled our dedicated filing team, ensuring our timeline for the pilot was reasonable and feasible. There were several considerations and planning steps involved:

START OF THE JOURNEY AND UPSKILLING OF THE TEAM

- Choosing the right study was an important consideration, we opted for a study that had a straightforward study design, standard analyses so there was familiarity with the programming specifications, and the reporting would align with existing Roche data standards. This would ease the application of the available open-source packages. In addition, considering the study timelines and ensuring that there was enough time to allow the team to enhance their proficiency with the new tools.

- Upskilling was, naturally, the most intense part of the filing team's journey. Forming a data science team who are motivated and passionate to upskill, was key to our success. Being an early adopter meant there was a lot to learn and not everything was perfect; the data science team members were not just end users but would co-create with the developers to shape our future system and tools. This mindset and seeing the bigger picture and wanting to be part of this transformative process was also key.

  The starting point for upskilling varied since we had a mix of R and SAS ® (but mostly SAS ®) programmers within the team and with varying clinical trial experience. The upskilling could cover basic R training, package specific training, git, and in addition, system related training, and learning new processes and ways of working on OCEAN.

  Allowing enough time to practise new tools and solutions on real clinical trial work is essential, as this is when the learning is embedded. Having a mix of good R programmers and those with clinical trial experience is crucial as it allows the team to work as 'one' and support each other in their respective learning journey.

PROGRAMMING STRATEGY

Given this was our first time using our NextGen Tools and System on a filing, ensuring an on-time completion was critical. To manage potential risks, whether from the health authority or internal challenges, we made a thoughtful programming strategy to cover the possibility that the filing might happen with the traditional SAS ® way. Therefore, the first line programming used R with the cutting-edge system and tools; the second line QC used - SAS ®. This dual approach also gave us confidence in the accuracy of our analyses, and provided further flexibility should we need the SAS ® version of the program in the future.

STAKEHOLDER MANAGEMENT & HEALTH AUTHORITY COLLABORATIONS

After aligning internally within our team, we needed to secure endorsement from the Global Filing Team and Global Development Team. Initially, we were questioned about the necessity, as the Global Team just wanted to have a smooth and quick filing. To convince them, we collaborated internally with the NextGen Team, to provide insights into trends in technology in the industry and the vision of our function; we also presented our risk management plan, highlighting that plan B, involving traditional methods, was ready as backup to ensure a timely filing. Building trust within the filing team was important. And later when we shared outputs produced using R, which looked slightly different to what they traditionally would expect to see, stakeholders were patient and supportive.

Another consideration was early communication with health authorities about the R-based submission to foster a collaborative and smooth submission process.

LEARNING EXPERIENCE AND CHALLENGES

Overall, our learning experience has been a positive journey, though it involved a steep learning curve. There was a lot to grasp - a new system, a new environment, and new tools and processes. The learning challenges are very individual. However, based on our team experience, getting used to how we work on OCEAN, using GitLab for version control (which impacted our QC workflow), and orchestrating analyses with SnakeMake, an open-source Python-based tool to publishing analyses to our document management system, were some of most challenging areas for users. Over time, after many fixes, system improvements, practice, we have noticed a growing sense of confidence and familiarity with these new tools. In contrast, training on Pharmaverse packages was easier to pick up, users liked the structured modular approach to programming, as it helps navigate and debug code more easily. The documentation is easy to follow, and template programs provide a consistent starting point, and are detailed enough for data scientists new to pharma to be able to pick up.

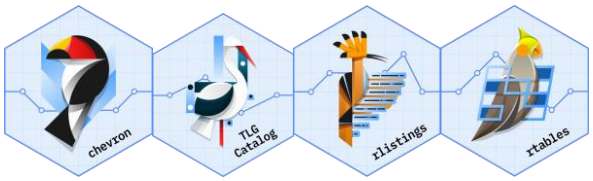
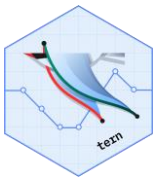
PROGRAMMING APPROACH

The programming strategy was to conduct all first-line programming in R, and QC using SAS ®. Table 1 below outlines the key R packages used for the preparation of SDTM, ADaM datasets and the creation of TLGs for the dossier. All the packages except for the Roche's version of {sdtm.roak} and the proprietary Roche {admiralroche] package are well-established, cross-collaboration solutions developed through Pharmaverse. The {admiralroche} package is based on all the functions from the {admiral} open-source family of packages, only with Roche-specific choices around ADaM implementation adhering to Roche's data standards.

This approach aimed to maximise the use of open-source packages to enhance transparency during the health authority review process. By leveraging these openly available packages, we provided reviewers with the opportunity to directly access source code, enabling them to address any questions or concerns more efficiently, should they wish to explore the underlying methods themselves.

Table 1. Key R packages used in the R submission.

Deliverables	**Key R packages:**
SDTM	{roak} package, Roche's version of now open-source {sdtm.oak} package: 

ADaM (inc Readable code)	Pharmaverse:  Proprietary: 
TLGs	Pharmaverse: 
Readable code for TLGs	Pharmaverse: 

THE REGULATORY SUBMISSION JOURNEY

The eSubmission was a global submission, to the FDA for submission in the U.S, with the EMA for submission in the EU, and the NMPA for submission in China. This was our first time preparing for an R-based submission, and our goal was to avoid any surprises and ensure a smooth process by proactively informing health authorities of our intentions. Beyond the technical preparation, being clear and having ongoing communication with health authorities proved to be a critical element in making our plan successful. Establishing this dialogue early allowed us to address any potential concerns and align expectations, which was essential for the overall success of the submission.

HEALTH AUTHORITY INTERACTIONS

FDA

For the U.S submission, we initiated communication with the FDA during the submission preparation period for the content and format, proposing the use of R and open-source packages as the readable code in the briefing package. In response, the FDA mentioned SAS® programs. After seeking clarification regarding FDA Statistical Software Clarifying Statement and the acceptance of the dummy R submission pilot from R consortium, we received a positive response from FDA reviewing team. FDA permits us to proceed with R, along with an appropriate justification of the R package used as well as any resources that provide evidence of validation of the package or function being used. This positive response allowed us to confidently continue our journey with R and we greatly appreciated the flexibility shown by the agency's reviewing team.

Shortly after our submission, we received a request for an application orientation meeting and a technical walkthrough. In addition to the usual content covering datasets and supporting documents in module 5, we included a list of the R related documents and their locations, such as the programTOC file and validation reports for the R packages. To help ease some concerns during the review, we also pointed out the differences in default options between R and SAS for the primary endpoint analysis.

We did not receive any information requests related to the R submission. Furthermore, we are excited to share that we have got the approval from the FDA – 7 weeks ahead of the FDA goal date.

EMA

For the submission to the EMA, the same eSub package was sent as for the FDA. We were fortunate to align with the application window for the EU Raw Data pilot, and the EMA promptly confirmed their acceptance. During the EU Raw Data pilot meeting, we presented content like the FDA technical walkthrough, but included additional details regarding the dataset content, and how they linked back to the analyses and explanations for the submitted documents, as it was the first time we submitted data to the EMA.

So far, there are no technical related questions from the EMA. The Raw Data Pilot report we received along with Rapporteur Day 80 report states that 'the replicated analysis is consistent with results provided by the applicant'.

NMPA

For the submission in China, initially, we raised this topic at a similar time as the FDA briefing package, but we did not get a clear response if the NMPA accepted the R submission or not. To address this, Roche requested a Type III meeting, providing more comprehensive supportive information in the guidance from health authorities, as well as the industry's usage of R in the submissions, and our usage of a validated R computing environment. At the beginning of this year, China CDE provided a written response stating that the biostatistics viewpoint is consistent with FDA, and there is no specific software requirement for analyses. Like the FDA, the response mentioned that the analysis software and corresponding functions used should be validated. We utilised the same eSub package submitted to the FDA and the EMA as the basis for the submission in China.

ELECTRONIC SUBMISSION PACKAGE

The data content and formats are consistent with CDISC requirements following industry-wide standards. Table 2. below outlines the eSub package content used in the R submission. Compared with a traditional eSub package the content remains the same.

The submitted datasets are all in .xpt format, however, generated using R via {xportr} package. The programTOC is a separate document detailing the readable codes included in the eSub package, which may also typically be included in the ADRG. The programTOC is normally included in a traditional submission package. The readable code for ADaM as detailed above uses {admiral}, {admiralonco} packages and Roche's proprietary {admiralroche} package, and for key outputs the readable code is generated using the {tern} package. As {admiralroche} contains the analysing standards for the ADaM datasets, we need to include this in the eSubmission to allow reviewers to easily replicate our results.

Table 2. eSub Contents for an end-to-end R study

	Content of Tabulation Package	**Content of Analysis Package**	**Format**
Pivotal study	• SDTM XPT • aCRF • Define.xml • cSDRG	• ADaM XPT • Define.xml • ADRG • Readable code for ADaM and Key outputs • programTOC	CDISC

FILE OF programTOC DISCLOSES R PACKAGES

FDA had requested to specify the details of the R packages used and justification of its use. This extra information was included in the programTOC file.

The programTOC file will generally consists of

- Description of the eSub package
- List of Analysis Dataset Programs
- List of Output Programs (list names of key analysis variables and datasets)
- How to use the readable code. In this section, we added the additional information requested by the FDA see the below extract in Table 3. which discloses the use of all the R packages used, the package version, and brief description detailing its use.

Table 3. Additional details submitted for the R submission in the programTOC file

Proprietary R Analysis Package	Package version	Description
admiralroche	0.3.0	Roche-specific extension of admiral. Provides Roche-specific functionality to facilitate ADaM programming using admiral.

R Packages	Package version	Description
admiral	0.12.1	A toolbox for programming Clinical Data Interchange Standards Consortium (CDISC) compliant Analysis Data Model (ADaM) datasets in R.Analysis derivations are implemented in accordance with the "Analysis Data Model Implementation Guide" (CDISC Analysis Data Model Team, 2021, <https://www.cdisc.org/standards/foundational/adam/adamig-v1-3-release-package>).
admiraldev	0.5.0	Utility functions to check data, variables and conditions for functions used in 'admiral' and 'admiral' extension packages.

We used the {pkglite} package to convert {admiralroche} package to a txt file which was then included with the other readable code files in the eSub.

The appendix of the programTOC file, contains detailed instructions on how to execute analysis programs in R to help reviewers. This included instructions on how to install the R environment, how to define a working directory, how to specify R package repository, how to install R packages, along with an example of executing the submitted readable code. The instruction to install {admiralroche} package using the .txt file at reviewers' end is also illustrated.

INTERNAL VALIDATION PROCESS

To meet FDA requests regarding the inclusion of validation reports for the R packages or functions used, we included validation reports for key R packages as part of the miscellaneous datasets. These reports are derived from our highly automated internal validation process. For more information, please refer to the recording mentioned below under Recommended Reading.

CONCLUSION

Looking back at our submission journey, it has been both challenging and valuable. Throughout the process of our new drug application submission using R, we have found that health authorities are open to R-based submissions, and fostering collaborative open two-way communication is key. To facilitate a smooth submission review, it's crucial to secure agreements of your intention before the submission of the dossier. From a practical standpoint, there was limited extra effort required in preparing the eSubmission package.

We will continue this journey and believe that the end-to-end R submissions in the future will be more efficient with the route that we paved from this experience. We hope that sharing our journey serves as encouragement and contributes to broader industry adoption to ultimately help deliver medicines faster.

ACKNOWLEDGMENTS

Many thanks to all the data scientists who upskilled and made this journey possible, as well as to the internal stakeholders (the Global Filing Team) for their support and partnership. The R Submissions journey has been about collaborations involving many functions. While the data science team drives this innovation, adaptability of other functions has been crucial to our success. We also appreciate the reviewing teams from health authorities, who have shown openness to our innovative proposals, enabling us to reach our goal. This has been a pioneering experience, made achievable only through the cross-industry collaborations, dedication and commitment of many people!

RECOMMENDED READING

- Shifting to an Open-Source Backbone in Clinical Trials with Roche
- {admiral} - from open source through to company implementation
- Roche's end-to-end R Journey to Submission
- A Risk-based Approach for Assessing R package Accuracy within a Validated Infrastructure
- R/Pharma 2022 Day 2: Coline Zeballos & Doug Kelkhoff. Roche's approach to software validation

CONTACT INFORMATION

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

Hinal Patel
Roche Hexagon Place,
Shire Park,
Falcon Way
Welwyn Garden City
AL7 1TW
Email: hinal.patel@roche.com

Jingyuan Chen
Genentech,
1 DNA Way
South San Francisco
94080, California, US
Email: chen.jingyuan@gene.com