PharmaSUG 2026
Paper for Presentation 163

Current Review of Open Source in New Drug Applications: R & Python

Phil Bowsher, Posit/RStudio PBC; Boston, USA,

Abstract

I started a list of submissions that use open source in 2023. I maintain the list at:

a. https://github.com/philbowsher/Open-Source-in-New-Drug-Applications-NDAs-FDA

The 2026 analysis documents the increased prevalence of R within New Drug Applications and its integration into the submission process, while Python appears with lower frequency. The current landscape indicates that open-source software is utilized as a primary tool for both Sponsor analytic submissions and FDA independent reviewer assessments.

Introduction

1. Key Takeaways: Open-Source-in-New-Drug-Applications-NDAs-FDA

A. 'Clinical Data Scientist': A formalized role within the Agency dedicated to R-based exploration and independent safety analysis.
https://www.accessdata.fda.gov/drugsatfda_docs/nda/2026/761458Orig1s000MultidisciplineR.pdf
2026 GlaxoSmithKline "The clinical data scientist performed an independent safety analysis using R Studio version 4.1.0. " "Source: adex.xpt and adsl.xpt; Software: R "

B. Regulatory Acceptance of Open Source Environments
https://www.accessdata.fda.gov/drugsatfda_docs/nda/2025/125504Orig1s063.pdf
2025 Novartis Pharmaceuticals Corporation. "For dataset creation, data manipulation, data presentation, construction of plots and logistic regression analyses, R version 3.6.2 and/or version 4.1.3 (the R Foundation for Statistical Computing) was used. Acceptable"

https://www.accessdata.fda.gov/drugsatfda_docs/nda/2025/761352Orig1s000MultidisciplineR.pdf
2025 Merus N.V. "R (Version 4.2.1, The R foundation for Statistical Computing). Rstudio (version 2022.07.1, RStudio Inc, Boston, USA)." "Acceptability [FDA's comments] Acceptable"

https://www.accessdata.fda.gov/drugsatfda_docs/nda/2025/219042Orig1s000AdminCorres.pdf
2025 Boehringer Ingelheim Pharmaceuticals, Inc. "In regard to FDA's comments #22a-c, Boehringer Ingelheim followed....csv file for datasets, .pdf file for define file, and .mod or .R or .lst file for modeling files." "Question 22a – c: FDA stated that the data format was generally acceptable."

C. OCS Analysis Studio used by FDA in reviews.
Built with R, Shiny, JavaScript, D3.js
https://www.fda.gov/media/171317/download

D. R's popularity in PK
https://www.accessdata.fda.gov/drugsatfda_docs/nda/2024/218213Orig1s001.pdf

2024 Bristol Myers Squibb Company "R (Version 3.6.1 or later); Xpose and PsN (Department of Pharmacy, Uppsala University, Uppsala, Sweden); Pirana (Version 2.9.1); R (Version 3.6.3 or above; R Foundation for Statistical Computing, Vienna, Austria); RStudio (Version 1.2.5033; RStudio, Boston, MA, USA)" "Yes, FDA agrees that the software and their versions used to conduct the analyses are acceptable. Standard software were used for conducting the PopPK analysis."

https://www.accessdata.fda.gov/drugsatfda_docs/nda/2024/218944Orig1s000MultidisciplineR.pdf
2024 Syndax Pharmaceuticals, Inc. "The statistical reviewer's analyses were performed using R 4.0.5 (The R Foundation, https://www.rproject.org/)." "R (Version 4.0.0 or higher) for simulations, graphs, and data manipulations."

E. Use of RMD
https://www.accessdata.fda.gov/drugsatfda_docs/nda/2025/219876Orig1s000MultidisciplineR.pdf
2025 Chimerix, Inc. "Source: CHIM-ONC201_diagnostics_run315.Rmd""Source: Chimerix_ONC201_ForestPlot_2024.R"

F. Large Pharmas
https://www.accessdata.fda.gov/drugsatfda_docs/nda/2026/761458Orig1s000MultidisciplineR.pdf
2026 GlaxoSmithKline "The clinical data scientist performed an independent safety analysis using R Studio version 4.1.0. " "Source: adex.xpt and adsl.xpt; Software: R "

https://www.accessdata.fda.gov/drugsatfda_docs/nda/2025/219208Orig1s000IntegratedR.pdf
2025 Novartis Pharmaceuticals "Source: Reviewer's analysis based on adsl.xpt; Software: R"

https://www.accessdata.fda.gov/drugsatfda_docs/nda/2025/761400Orig1s000MultidisciplineR.pdf 2025 Regeneron Pharmaceuticals, Inc. "and Rstudio"

G. Broad Use
https://www.accessdata.fda.gov/drugsatfda_docs/nda/2024/218944Orig1s000MultidisciplineR.pdf
2024 Syndax Pharmaceuticals, Inc. "The statistical reviewer's analyses were performed using R 4.0.5 (The R Foundation, https://www.rproject.org/)." "R (Version 4.0.0 or higher) for simulations, graphs, and data manipulations."

https://www.accessdata.fda.gov/drugsatfda_docs/nda/2024/213586Orig1s000MultidisciplineR.pdf
2024 Teva Neuroscience Inc. "Primary analysis and plotting were performed in R 4.1.2. "

H. Roche Complete Open Source Analytic Pipeline and Compound Acceptance
https://www.accessdata.fda.gov/drugsatfda_docs/nda/2024/219249Orig1s000MultidisciplineR.pdf
2024 Genentech Inc. "were performed using R® V4.2 (or higher) with comprehensive R archive network (CRAN)"
https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-inavolisib-palbociclib-and-fulvestrant-endocrine-resistant-pik3ca-mutated-hr-positive
https://www.youtube.com/watch?v=4YpKAaI5WUU
https://phuse.s3.eu-central-1.amazonaws.com/Archive/2024/Connect/EU/Strasbourg/PAP_SA02.pdf
https://phuse.s3.eu-central-1.amazonaws.com/Archive/2024/Connect/EU/Strasbourg/PRE_SA02.pdf
https://www.gene.com/media/press-releases/15011/2023-12-04/genentech-announces-positive-phase-iii-r

I. Some Python
https://www.accessdata.fda.gov/drugsatfda_docs/nda/2025/219792Orig1s000IntegratedR.pdf
2025 UCB INC "Source: adsl.xpt; Software: R" "and Python software"

https://www.accessdata.fda.gov/drugsatfda_docs/nda/2025/022137Orig1s015MultidisciplineR.pdf
2025 022137 Label Change "identified through a python-based tool"

J. R Software Version Variance
https://www.accessdata.fda.gov/drugsatfda_docs/nda/2026/761458Orig1s000MultidisciplineR.pdf
2026 GlaxoSmithKline "The clinical data scientist performed an independent safety analysis using R Studio version 4.1.0. " "Source: adex.xpt and adsl.xpt; Software: R "

https://www.accessdata.fda.gov/drugsatfda_docs/nda/2025/219839Orig1s000MultidisciplineR.pdf
2025 Dizal (Jiangsu) Pharmaceutical Co Ltd. "R (Version 4.2.3)"

https://www.accessdata.fda.gov/drugsatfda_docs/nda/2025/125504Orig1s063.pdf
2025 Novartis Pharmaceuticals Corporation. "For dataset creation, data manipulation, data presentation, construction of plots and logistic regression analyses, R version 3.6.2 and/or version 4.1.3 (the R Foundation for Statistical Computing) was used. Acceptable"

K. Package Source and Integrity: Documentation indicates Agency acceptance of the Comprehensive R Archive Network (CRAN) as a standard repository for analysis packages.
https://www.accessdata.fda.gov/drugsatfda_docs/nda/2024/219249Orig1s000MultidisciplineR.pdf
2024 Genentech Inc. "were performed using R® V4.2 (or higher) with comprehensive R archive network (CRAN)"
https://cran.r-project.org/web/packages/adaptIVPT/index.html 2024-01-26
Package provides procedures based on parallel replicate design and balanced data, according to the U.S. Food and Drug Administration's "Draft Guidance on Acyclovir"
https://www.accessdata.fda.gov/drugsatfda_docs/psg/Acyclovir_topical%20cream_RLD%2021478_RV12-16.pdf
Maintained by FDA

Conclusion

This paper highlights R's use in submissions in NDAs by both sponsors and reviewers. These are current trends applicable for statistical programmers. The information outlined in this paper highlights the evolving space of using open source for submissions.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:
Phil Bowsher
phil@posit.co