Paper OS01

How R We Now? Tracking Industry Progress in Adopting R and Open-Source for Clinical Reporting

Andrzej Białaś, Appsilon, Warsaw, Poland
Ashley Tarasiewicz, Atorus Research, Philadelphia, PA, USA

ABSTRACT

This paper summarizes results from the **2025 Pharmaverse R Adoption Survey**, the second annual snapshot of how sponsor companies use **R and open-source tooling** across clinical reporting (CRF → submission). This year's survey builds upon our initial assessment conducted nearly 1.5 years ago, allowing for a direct comparison of results and providing insights into how adoption rates, practices, and perceptions have evolved over time. The 2025 edition also adds questions on **Python usage** in clinical reporting.

By presenting these updated findings, we aim to highlight key trends, identify ongoing challenges, and share lessons learned. In addition, we will discuss our continued efforts and experiences in fostering open-source collaborations across the pharmaceutical industry to accelerate innovation and improve clinical reporting practices.

The survey also highlights what slows adoption down, primarily **resistance to change** and **package/dependency management.** At the same time we clearly see what helps it scale – **training**, **internal champions**, **executive sponsorship**, and a growing body of **regulatory precedent**.

A quick note: the numbers below describe where organizations *say* they are today. They don't prove maturity. But they do show direction which matters when planning standards, validation, training, and tooling.

INTRODUCTION

Pharmaverse exists to support collaborative development of open-source tools for clinical reporting. The goal is simple: reduce duplicated effort, improve consistency, and make regulated adoption less painful.

In 2024, Pharmaverse ran its first adoption survey to "pulse check" how far sponsor companies had progressed, and what they planned next. In that first survey, **20 out of 23** companies expected R usage to increase over the coming years (with only **one** predicting a decrease).

In 2025, we repeated the survey to track change over time and broaden the view to include Python-based approaches alongside R. We received **31 responses** after removing entries that were out of scope for this analysis.

Most respondents expect R usage to increase over the next two years, but adoption still looks uneven by task area. R use is most established in **TLG/TLF work** and **interactive exploration**, while **SDTM** remains earlier-stage for many teams. And while submission work is still not "everyday practice," it is no longer a distant idea — open-text responses describe real timelines, pilots, and internal readiness programs.

SURVEY DESIGN AND SAMPLE

THE QUESTION DESIGN

The 2025 survey followed closely on the structure and design of last year's survey. The full list of questions can be found in **Appendix A** at the end of this document.

The survey included three types of questions: **open text**, **single choice** and **multiple choice**. While we had a total of 31 responses, not every answer count will sum up to 31 because respondents could select multiple answers for some questions. For multiple choice questions the totals reflect the number of answer selections, not number of companies. The answers to open text questions were reviewed by the authors of this paper and grouped into common themes where possible, and served to inform the commentary on the results.

The full survey can be accessed for viewing purposes, but as of 22nd of January 2026 is no longer accepting responses.

SURVEY TARGETS

The survey targets **pharma and biotech companies** and is not aimed at CROs or vendors, so any answers we received that were not from survey target companies were omitted in the analysis.

RESPONSES INCLUDED IN THIS ANALYSIS

This 2025 analysis includes **31 responses**, after excluding entries treated as out of scope for this report. We combined responses gathered through multiple collection routes. A subset of responses was captured outside Google Forms due to access constraints, then mapped into the same question structure so results could be analyzed together.

WHERE R ADOPTION STANDS IN 2025

THE NEAR TERM TREND IS CLEAR: R IS MOVING UP

Of the **31** company responses, **28** reported **increasing** R usage over the next two years and **3** reported R usage as **stable**. That's broadly consistent with last year's direction-of-travel: in 2024, **20 out of 23** also reported increasing R usage.

Most of the companies that reported increasing R usage shared comments that they are planning and preparing for steady year over year growth in this area and/or open-source becoming the primary driver in clinical reporting by 2030.

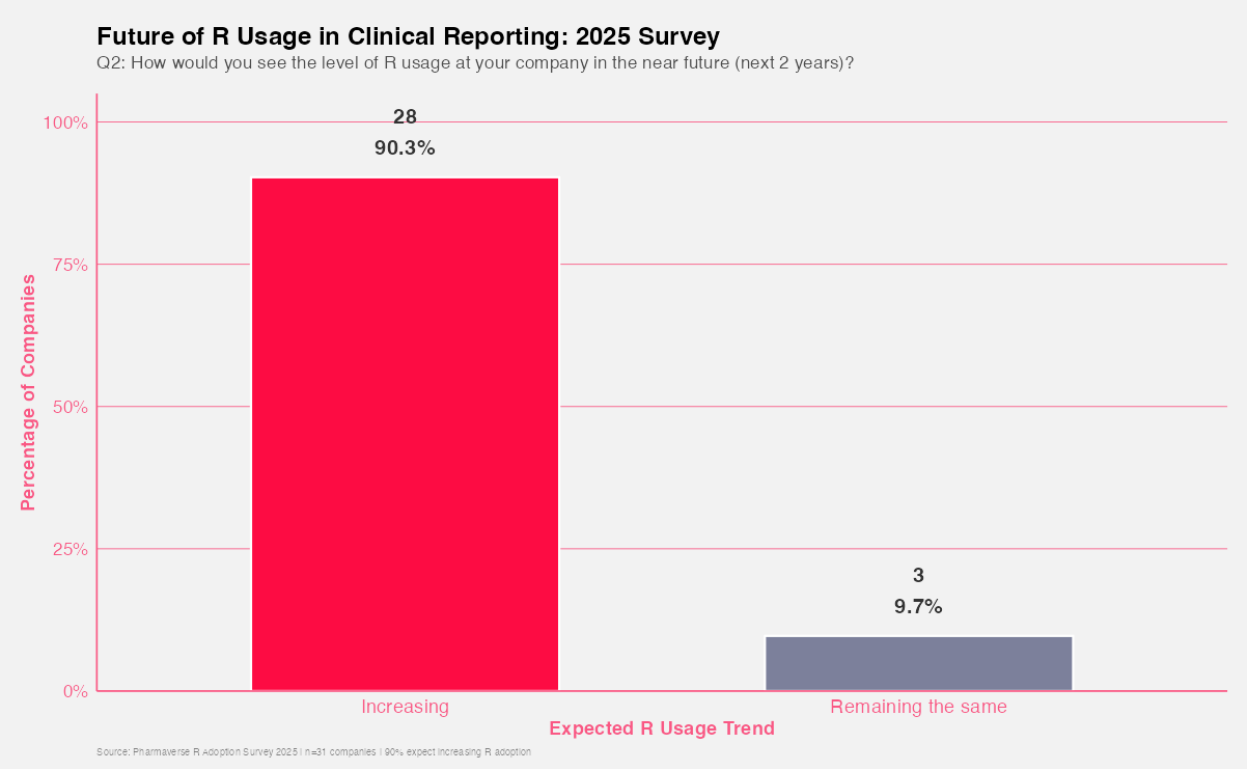

Figure 1: Future of R Usage Results

TEAM SIZES ARE UNEVEN

Of the **31** responses:

- **16** estimate **<50** R users doing clinical reporting tasks
- **8** estimate **50–249**
- **3** estimate **250–500**
- **4** estimate **>500**

So the "typical" company in this sample is still running R with a relatively small group but a meaningful minority is already operating at scale.

ADOPTION BY TASK AREA: ADAM AND TLG/TLFS ARE AHEAD; SDTM LAGS

A simple way to read the 2025 results is: **QC is mainstream**, first-line is growing, and SDTM is still catching up.

SDTM:

- **14 out of 31**: not currently, but plan to in future
- **10 out of 31**: QC
- **7 out of 31**: exploratory
- **2 out of 31**: first-line programming
- **6 out of 31**: no plans

Last year's SDTM baseline was lower for QC: **3** companies reported QC use in SDTM in 2024, with **8** exploratory and **2** first-line. So SDTM looks like it's shifting from "explore" toward "control and checks," which is often how regulated change starts. Nearly half (14 out of 31) of responses note no current use but planned for the future.

ADAM:

- **20 out of 31**: QC
- **14 out of 31**: exploratory
- **10 out of 31**: first-line programming
- **7 out of 31**: not currently, but plan to in future
- **3 out of 31**: no plans

In 2024, ADaM was already ahead: **13** exploratory, **10** QC, **7** first-line. What changed in 2025 is the scale of QC (now **20**). That's not a small shift. It suggests ADaM is moving from "pockets of use" into something closer to an operating model.

TLG/TLFS:

- **18 out of 31**: QC
- **17 out of 31**: exploratory
- **16 out of 31**: first-line programming
- **9 out of 31**: not currently, but plan to in future
- **1 out of 31**: no plans

In 2024, TFL usage was also broad: **19** exploratory, **16** QC, **13** first-line. So the 2025 story here is more "steady expansion" than "breakthrough."

INTERACTIVE DATA DISPLAYS (E.G. SHINY):

- **24 out of 31**: exploratory
- **10 out of 31**: QC
- **8 out of 31**: first-line programming
- **3 out of 31**: plan to in future
- **2 out of 31**: no plans

In 2024, Shiny usage was already high: **21** exploratory, **8** QC, **5** first-line. 2025 continues that trajectory, and it reinforces a familiar pattern: interactive tooling is often the gateway, but it doesn't stay "just exploratory" for long.

2025 CLINCAL REPORTING AREAS SUMMARY:

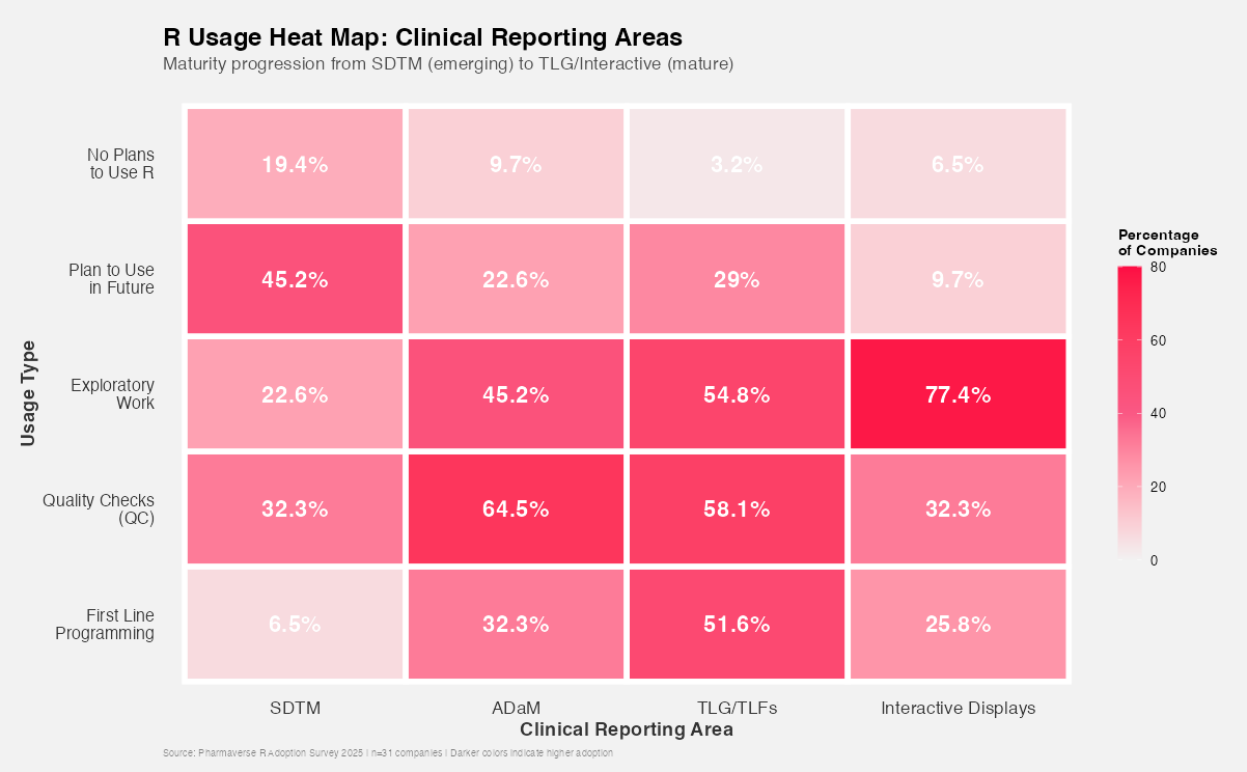

Figure 2: R Usage Heat Map for Clinical Reporting Areas

ELECTRONIC SUBMISSION:

- **3 out of 31**: doing end-to-end R-based eSubmissions
- **5 out of 31**: no plans
- **9 out of 31**: use R for some eSubmission components
- **16 out of 31**: not currently, but plan to in future

In 2024, the split looked similar: **12** reported "not currently, but plan to in future," **7** reported using R for some eSubmission components, **4** reported no plans, and **3** reported doing end-to-end R-based eSubmissions.

PYTHON USAGE

This year's survey added questions related to usage of Python in clinical reporting. Understanding that this is an emerging area with low adoption, there were only two questions, asking respondents to describe shortly the usage of Python in clinical reporting. Nineteen respondents provided a written answer about Python usage in clinical reporting tasks.

While Python is without a doubt showing up (most of the respondents noted some use of Python in their work), it is mostly used for exploratory work, pilot projects, experimental adoption or automation, rather than the core regulatory submissions.

What we saw most often in the replies:

- No or minimal usage (several responses were essentially "none," "a little," or "exploratory")
- Automation/orchestration/pipeline work (utilities, process tooling)
- ML/AI-adjacent work (modeling workflows, AI integration, experimentation)

- Web apps and APIs (in a few cases)

This was also reflected in the follow-up question asking for pharmaverse Python solutions used in this area. All respondents but one reported they don't use neither py-pkglite nor rtflite.

WHAT TEAMS REPORT GETTING OUT OF R

On "business impact observed from R adoption in clinical reporting", the most commonly reported outcomes were:

- **Increased analytical capabilities (23 out of 31)**
- **Cost savings vs. legacy tools (19 out of 31)**
- **Better collaboration across teams (12 out of 31)**

A smaller group selected:

- **Enhanced reproducibility (7 out of 31)**
- **Reduced timelines for deliverables (7 out of 31)**
- **Improved quality/consistency of outputs (4 out of 31)**
- **No measurable impact yet / too early to assess (3 out of 31)**

That last point matters. There's still a "prove it" phase happening in parts of the industry, even with adoption rising.

When reporting on the impact of adopting open-source solutions and tooling most of the respondents highlight the practical side of it: better analytical capability, lower costs and better, faster collaboration across teams. A small group reports it's too early to measure results.

WHAT SLOWS ADOPTION DOWN (AND WHAT HELP IT SCALE)

The biggest obstacles are within people and processes with the majority of respondents reporting resistance to change from existing workflows as the primary blocker. A practical headache of maintaining packages and dependencies over time is following up closely, as well as skill/training gaps. **Overall, the results suggest most organizations are past the "can we use R here?" stage and are now dealing with operationalizing R at scale**. Change management looks to be mixed rather than uniform. The majority of organizations describe adoption as project-specific implementations, with both top-down mandates and bottom-up "grassroots" adoptions equally common.

For the organizations that deem their R adoption as successful, the key reported success factors stemmed from committed leadership and strong internal champions backed by capability building. Respondents also pointed out that robust IT infrastructure and access to validated packages make adoption realistic in day-to-day work.

Top main obstacles identified in this year's survey:

- **Resistance to change from existing workflows (19)**
- **Package maintenance and dependency management (15)**
- **Lack of internal R expertise/training (12)**
- **Regulatory concerns/validation requirements (12)**
- **Quality control and reproducibility concerns (11)**
- **Integration with existing systems (9)** and **IT security/infrastructure constraints (9)**

Top key success factors:

- **Strong internal R champions (15)**
- **Executive sponsorship and leadership support (14)**
- **Dedicated training and upskilling programs (13)**
- **Regulatory precedent from other companies (11)**
- **Robust IT infrastructure setup (10)**
- **Use of pharmaverse resources (9)**
- **Demonstrable efficiency gains/ROI (8)**

While almost all companies reported blockers, the primary obstacle was mostly organizational and is identified as change management. The package layer is also a real challenge (maintenance and dependency management was selected **15** times) but the biggest issue is still change.

PACKAGE USAGE REFLECTS MATURITY (AND WHERE GAPS ARE)

Package usage lines up with the "ADaM ahead, SDTM catching up" picture.

- In **ADaM**, **admiral** (and extensions) is selected by **21** companies; **metatools (8)** and **metacore (7)** follow.
- In **SDTM**, **19** selected **N/A** for pharmaverse packages in that area; usage is spread across **metatools (7)**, **metacore (6)**, **sdtm.oak (6)**, **sdtmchecks (6)**, with a smaller group reporting **datacutr (4)** and **aNCA (1)**.

For TLG/TLFs, adoption is broad and the tooling ecosystem is diverse:

- Most used include **ggsurvfit (15)**, **rtables (15)**, **teal (13)**, and **tern (13)**.

And in "other packages," we see a pattern that matches the operational story:

- **diffdf (12)**, **logrx (11)**, **pharmaverseadam (10)**, **pharmaversesdtm (9)**, **riskmetric (6)**, and **datasetjson (6)** are all widely selected. Other than those, the respondents reported: whirl (3), pharmaverseraw (3), envsetup (2), covtracer (2), risk.assess (2), connector (1), riskassessment (1) and thevalidatoR (1).

WHAT RESPONDENTS WANT NEXT (FROM OPEN-TEXT RESPONSES)

Fifteen respondents shared feedback on gaps or collaboration ideas. Many said "no gaps," but here we share select, detailed comments clustered around a few themes:

- **Submission-ready environments.** There's an appetite for "can we run R in a way that satisfies internal and regulatory expectations" and not just "can we use R here?"
- **Metadata handling** has been a common theme in several comments.
- **Discoverability and documentation.** A lot of the respondents are overwhelmed with the number of available packages. Knowing what to use (and what to trust) is a common challenge.

What we note is that adoption is dependent on how the whole machine runs from standards, through validation approach, infrastructure, training, and governance.

A QUICK LOOK BACK: WHAT CHANGED SINCE 2024?

The 2024 paper captured a clear baseline across tasks—ADaM (13 exploratory / 10 QC / 7 first-line) and SDTM (8 exploratory / 3 QC / 2 first-line), plus broad TFL and Shiny usage.

It also showed early but real submission activity: **7** companies using R for some submission components, **3** doing end-to-end R submissions, and **12** planning to use R in submission work in the future.

Compared with that baseline, the 2025 results suggest:

- More weight moving into **QC** across domains (especially ADaM and SDTM).
- Continued growth in **interactive exploration**, with a gradual pull toward production use.
- A practical shift in mindset: from "what package do we use?" to "how do we validate, run, and govern this at scale?"
- A similar split in eSubmission as in 2024.

LIMITATIONS

It's important to understand that this is not a randomized study. It's a community survey, and the respondent mix can change year to year. So year-over-year comparisons should be read as directional, not definitive.

Also, “R usage” can mean different things depending on organization structure. A company might report “first-line usage” in one group and “no plans” in another and both can be true.

CONCLUSIONS AND NEXT STEPS

The 2025 survey reinforces last year’s main message: pharma and biotech companies continue to move toward R and open source in clinical reporting.

What feels different this year is the visibility of the middle. More teams seem past curiosity. They’re building real operating models and that surfaces the obstacles they encounter along the way even more: change management, governance, dependency control, and validation strategy.

If you’re trying to move adoption forward inside your organization, the data points to a practical path:

- Start where value is easiest to prove (often **QC**, utilities, and **TLG/TLF** work).
- Invest early in **training** and **internal champions**—they show up as top success factors.
- Treat **dependency management** and **reproducibility** as first-class work, not cleanup.
- Use **regulatory precedent** to reduce perceived risk; it’s repeatedly cited as a reason adoption becomes “allowed,” not just “possible.”

APPENDIX A:

A list of 2025 Pharmaverse Survey questions (without technical or feedback questions)

1. Estimated number of R users doing clinical reporting tasks for your company? [single choice]
2. How would you see the level of R usage for clinical reporting at your company in the near future (i.e. next 2 years)? [single choice]
3. Any additional comment on the previous question about trend of R usage for clinical reporting at your company in the near future? [open text answer]
4. Has your company set any targets relating to R usage, e.g. full R based eSubmission by YYYY, or GxP R readiness by YYYY? [open text answer]
5. What have been the main obstacles or challenges your organization has faced in adopting open source R for clinical reporting? (Select all that apply) [multiple choice]
6. For organizations that have successfully increased R adoption, what were the key success factors? (Select all that apply - skip if not applicable) [multiple choice]
7. What business impact has your organization observed from R adoption in clinical reporting? (Select all that apply) [multiple choice]
8. How would you characterize the change management approach for R adoption at your organization? (Select all that apply)
9. [SDTM] Please describe the R usage at your organisation with respect to the SDTM task. (select all that apply) [multiple choice]
10. Please select any of the following pharmaverse packages you use in this area. [multiple choice]
11. [ADaM] Please describe the R usage at your organisation with respect to the ADaM task. (select all that apply) [multiple choice]
12. Please select any of the following pharmaverse packages you use in this area. [multiple choice]
13. [TLG/TLFs] Please describe the R usage at your organisation with respect to the TLG/TLFs task. (select all that apply) [multiple choice]
14. Please select any of the following pharmaverse packages you use in this area. [multiple choice]
15. [Interactive Data Displays] Please describe the R usage at your organisation with respect to the Interactive Data Displays (e.g. Shiny) task. (select all that apply) [multiple choice]
16. Please select any of the following pharmaverse packages you use in this area. [multiple choice]

17. [eSubmission] Please describe the R usage at your organisation with respect to the eSubmission task. (select all that apply) [multiple choice]
18. Please select any of the following pharmaverse packages you use in this area. [multiple choice]
19. [Python] Please describe the Python usage at your organisation with respect to clinical reporting tasks. [open text answer]
20. Please select any of the following pharmaverse Python solutions you use in this area. [multiple choice]
21. [Other packages] Do you use any of the following further pharmaverse packages? [multiple choice]
22. Are there any other open source or internal R packages you use for clinical reporting tasks? And if any internal, would there be any future plans to open source them? [open text answer]

ACKNOWLEDGEMENTS

We would like to thank **Ari Siggaard Knoph** and **Ross Farrugia** for leading the 2024 survey and paper and for setting the foundation for this work. We also thank the **Pharmaverse Council** for their guidance and support in delivering the 2025 edition. Finally, **thank you to all the respondents** who took the time to represent their organizations and share practical input with us. Your contributions make this survey valuable to the whole community.

RECOMMENDED READING

1. Pharmaverse website and blog – where you can find more on any pharmaverse packages mentioned in this paper
2. PHUSE EU 2024 Survey paper
3. PHUSE EU 2024 Survey slides
4. Join the Pharmaverse Slack Community and start contributing to open source!

CONTACT INFORMATION

Contact the authors at:

Author Name: Andrzej Białaś
Company: Appsilon
Email: andrzej@appsilon.com
Website: appsilon.com

Author Name: Ashley Tarasiewicz
Company: Atorus Research
Email: ashley.tarasiewicz@atorusresearch.com
Website: atorusresearch.com