AMERICAN JOURNAL OF AI CYBER COMPUTING MANAGEMENT
PEER REVIEWED, REFERRED & INDEXED JOURNAL
E-ISSN:3069-0102
VOL.5, NO. 1(2025)

Enterprise-Scale Transition from SAS to R in Regulated Clinical Environments: Implementation Framework, Validation Strategy, and Lessons from a Global Pharmaceutical Deployment

Dharma Dev Bommi
Principal Statistical Programmer

Abstract—The pharmaceutical industry is at an inflection point in clinical statistical programming, with growing momentum to transition from proprietary SAS software to open-source R. While R offers compelling advantages in reproducibility, flexibility, and cost reduction, the shift in a GxP-regulated environment introduces substantial challenges: computer systems validation, regulatory acceptance, change management, and quality assurance at scale. This paper presents a structured implementation framework derived from a real-world enterprise-wide SAS-to-R migration executed at a leading global pharmaceutical company. The programme resulted in the design, validation, and deployment of CARP, a proprietary R-based clinical reporting platform built on more than 50 validated R functions, serving over 300 global users across five functional departments. Cumulative software licensing savings exceeded $50 million over an eight-year projection, while reporting turnaround times decreased by an average of 55 percent across key clinical deliverables. Quality metrics including reproducibility scores, error rates, and audit trail completeness showed material improvement post-migration. The paper documents the platform architecture, the Computer Systems Validation (CSV) approach adopted under 21 CFR Part 11, EU Annex 11, and GAMP5 guidelines, and the four-phase change management strategy that enabled organisation-wide adoption. The framework described herein is intended as a practical reference for statistical programming organisations contemplating a similar transition in regulated clinical environments.

Keywords—Clinical Statistical Programming, Open-Source R, SAS Migration, CDISC, GxP Validation, ADaM, SDTM, Computer Systems Validation, Regulatory Submissions, Clinical Data Automation, Digital Health Transformation

I. INTRODUCTION

Clinical statistical programming is the operational backbone of regulated drug development. Every clinical trial submitted to regulatory authorities such as the U.S. Food and Drug Administration (FDA), the European Medicines Agency (EMA), or the Pharmaceuticals and Medical Devices Agency (PMDA) depends on standardised clinical datasets and analysis outputs generated by statistical programmers following guidelines set by the Clinical Data Interchange Standards Consortium (CDISC) [1]. For more than three decades, SAS (Statistical Analysis System, SAS Institute Inc.) has been the *de facto* programming language for this work, deeply embedded in regulatory workflows, validated computing environments, and institutional expectations globally.

However, over the past decade, the open-source R ecosystem has matured substantially. The availability of purpose-built packages for clinical data analysis, including {admiral}, {haven}, {ggplot2}, {Hmisc}, and the broader tidyverse, combined with the absence of licensing costs relative to enterprise SAS, has created powerful economic and scientific incentives to transition. The FDAs 2015 Statistical Software Clarifying Statement explicitly confirmed that the agency does not require the use of any specific software package, opening the regulatory door to R-based submissions [2].

Subsequent FDA and EMA guidance documents on data standardisation have reinforced this tool-agnostic position, and the R Consortium Submissions Working Group has demonstrated R viability in live regulatory filings [7].

Despite these tailwinds, enterprise-scale adoption of R in regulated environments remains limited. The barriers are not primarily technical but institutional: computer systems validation (CSV) under 21 CFR Part 11 (United States), EU Annex 11 (Europe), and GAMP5 guidance requires extensive documented effort; change management across large cross-functional organisations is complex; and the perceived risk of regulatory queries during submission review creates inertia. As a result, most published accounts of R adoption describe pilot programmes, proof-of-concept work, or individual-team transitions rather than full enterprise deployments serving hundreds of users across multiple therapeutic areas and regulatory jurisdictions [6, 9].

This paper addresses that gap directly. We document the architecture, validation strategy, change management approach, and quantified outcomes of an enterprise-wide SAS-to-R migration at a major global pharmaceutical company, culminating in the deployment of the CARP platform (the organisation Clinical Analytics and Reporting Environment), a proprietary R-based clinical reporting system serving over 300 global users across statistical programming, biostatistics, data management, regulatory affairs, and medical affairs. CARP was built on more than 50 validated R functions, delivered regulatory submissions to the FDA, EMA, and PMDA, and generated cumulative licensing savings exceeding $50 million over an eight-year projection period.

The contributions of this paper are threefold. First, we provide a detailed technical description of an enterprise-grade R reporting architecture suitable for GxP-regulated environments. Second, we document a four-phase change management framework empirically derived from direct organisational deployment experience across 18 months. Third, we present quantified performance metrics measured before and after migration, spanning adoption rates, quality outcomes, cost savings, and regulatory acceptance signals. The remainder of the paper is structured as follows: Section II reviews relevant literature; Section III describes the case study context; Section IV presents the CARP implementation framework; Section V reports and discusses results; and Section VI concludes with lessons for the industry.

II. RELATED WORK

The use of open-source statistical tools in pharmaceutical drug development has attracted growing academic and industry attention since the early 2010s. Pioneering work by Roche, Novartis, and GSK established early proof of concept that R could be used in regulated environments for clinical data analysis, with findings disseminated primarily through PhUSE working group outputs and conference proceedings [3]. These early accounts established that R was technically capable of producing regulatory-grade outputs but lacked the formal validation governance frameworks needed for enterprise-scale adoption.

The FDAs 2015 Statistical Software Clarifying Statement was a watershed moment for the industry, explicitly stating that the agency does not recommend or endorse any particular software for statistical analyses, and that it is the responsibility of the sponsor to ensure the appropriateness of the chosen software [2]. Subsequent FDA guidance on the CDISC Study Data Technical Conformance Guide (2020) made no explicit software requirement, reinforcing this position [4]. The EMA Reflection Paper on Artificial Intelligence (2021) further signalled that regulatory expectations were evolving toward tool-agnostic frameworks, provided appropriate validation evidence was in place [5].

From a validation standpoint, GAMP5 (A Risk-Based Approach to Compliant GxP Computerised Systems, revised 2022) provides the principal industry framework for validating computerised systems used in regulated life sciences [11]. The R Validation Hub Working Group has applied GAMP5 principles specifically to open-source statistical software, proposing a risk-based categorisation approach that treats standard R packages as Category 3 (non-configurable software) and custom internal packages as Category 4 or 5 (configurable or bespoke software) depending on the degree of customisation [6]. This categorisation underpins the validation approach described in Section IV-B of this paper.

The R Consortium Submissions Working Group achieved a landmark milestone in 2021, completing the first successful electronic regulatory submission to the FDA using entirely R-generated outputs [7]. This submission demonstrated that R can meet FDA requirements for reproducibility, traceability, and data integrity in a regulatory context, significantly accelerating industry confidence in R adoption. The Pharmaverse consortium, an industry-wide collaboration, subsequently published the {admiral} package for ADaM dataset construction and the {metacore} package for metadata management, further standardising R-based clinical programming workflows. These community-driven developments have significantly reduced the barrier to validated R adoption for organisations that leverage them.

On change management, the Prosci ADKAR model (Awareness, Desire, Knowledge, Ability, Reinforcement) has been widely cited in healthcare IT transformation literature as effective for enterprise software transitions [8]. Gnanasambandapillai et al. (2021) examined lessons from a large-scale SAS-to-R migration programme in a regulated clinical environment, identifying leadership sponsorship and peer mentorship networks as the two most critical enablers of sustained adoption, ahead of formal training provision [9]. These findings are consistent with observations from the deployment described in this paper. What distinguishes our contribution from prior literature is its scope and empirical grounding: we report on a fully completed, regulatory-tested, enterprise-scale R deployment with quantified outcomes across adoption, quality, efficiency, and cost dimensions, a level of specificity largely absent from existing published accounts.

III. CASE STUDY CONTEXT AND DATA OVERVIEW

The case study organisation is a top-five global pharmaceutical company headquartered in Cambridge, United Kingdom, with clinical programming operations distributed across North America, Europe, and the Asia-Pacific region. At the initiation of the migration, the statistical programming function comprised approximately 320 staff globally, the majority of whom conducted clinical reporting work exclusively within a centrally managed SAS environment built on the Life Science Analytics Framework (LSAF). SAS was the only approved language for regulatory submissions, and the organisation had no established R validation infrastructure at project commencement.

The therapeutic portfolio spanned oncology, respiratory, diabetes, immunology, and rare disease programmes, with active regulatory submissions to the FDA, EMA, PMDA, and Health Canada at the time of migration initiation. All CDISC submission datasets (SDTM, ADaM) and analysis outputs (TLFs, define.xml, cSDRG, ADRG) were required to comply with applicable CDISC implementation guides and health authority conformance guides. Therapeutic area coverage included flagship programmes in respiratory medicine (notably the a flagship respiratory medicine programme global programme impacting approximately 250 million patients) and COVID-19 vaccine development, providing a rigorous real-world test of the platforms regulatory readiness.

Migration planning commenced in Q4 2021, with CARP platform development beginning in Q1 2022. Full deployment, defined as exceeding 80 percent active user adoption across all target functions, was achieved in Q4 2023, representing an 18-month total programme duration. The migration was executed in parallel with ongoing regulatory submissions, a significant operational constraint that necessitated careful phasing and risk management.

The performance datasets used for comparative analyses in this paper were drawn from four sources: (a) internal time-tracking and project management records for turnaround time analysis, covering 214 clinical deliverables pre-migration and 187 deliverables post-migration; (b) quality control audit logs from the Smart Validation Engine, covering 2,840 dataset validations across both environments; (c) platform analytics from the CARP adoption dashboard, tracking active user counts and module utilisation quarterly; and (d) organisational finance records for SAS licensing expenditure and R platform investment. All data have been aggregated or indexed to remove commercially sensitive values, and cost savings estimates are presented as cumulative eight-year projections based on observed annualised savings in years one and two.

TABLE I. Comparative Assessment: SAS vs R in Regulated Clinical Programming Environments

Criterion	SAS (Legacy Environment)	R / CARP Platform
Licensing Model	Proprietary; ~$8.2M/yr enterprise	Open-source (free); ~$0.8M/yr maintenance
Regulatory Acceptance	Universally accepted (FDA, EMA, PMDA)	Accepted; FDA/EMA confirmed tool-agnostic (2015+)
Reproducibility	72% full reproducibility across machines	96% via {renv} environment locking

Validation Framework	Established IQ/OQ/PQ protocols; vendor-supported	GAMP5-aligned CSV; custom V-model framework
Output Formats	RTF, PDF, SAS datasets (SAS7BDAT)	RTF, PDF, HTML, XPT, RDS, RMarkdown reports
Version Control	Limited; SAS code stored in LSAF folders	Full Git integration; branch/tag per submission
Scalability / Extensibility	Limited; extensions require SAS licensing	High; 15,000+ CRAN packages available
Mean Report Turnaround	21 days (SDTM); 28 days (ADaM)	9 days (SDTM); 11 days (ADaM) - 55% reduction

IV. THE CARP FRAMEWORK: PROPOSED METHODOLOGY

A. Platform Architecture

CARP follows a five-layer architecture designed to replicate and materially extend the clinical reporting capabilities of the legacy SAS LSAF environment, while meeting GxP compliance requirements across the full regulatory submission lifecycle. Figure 1 illustrates the complete architecture.

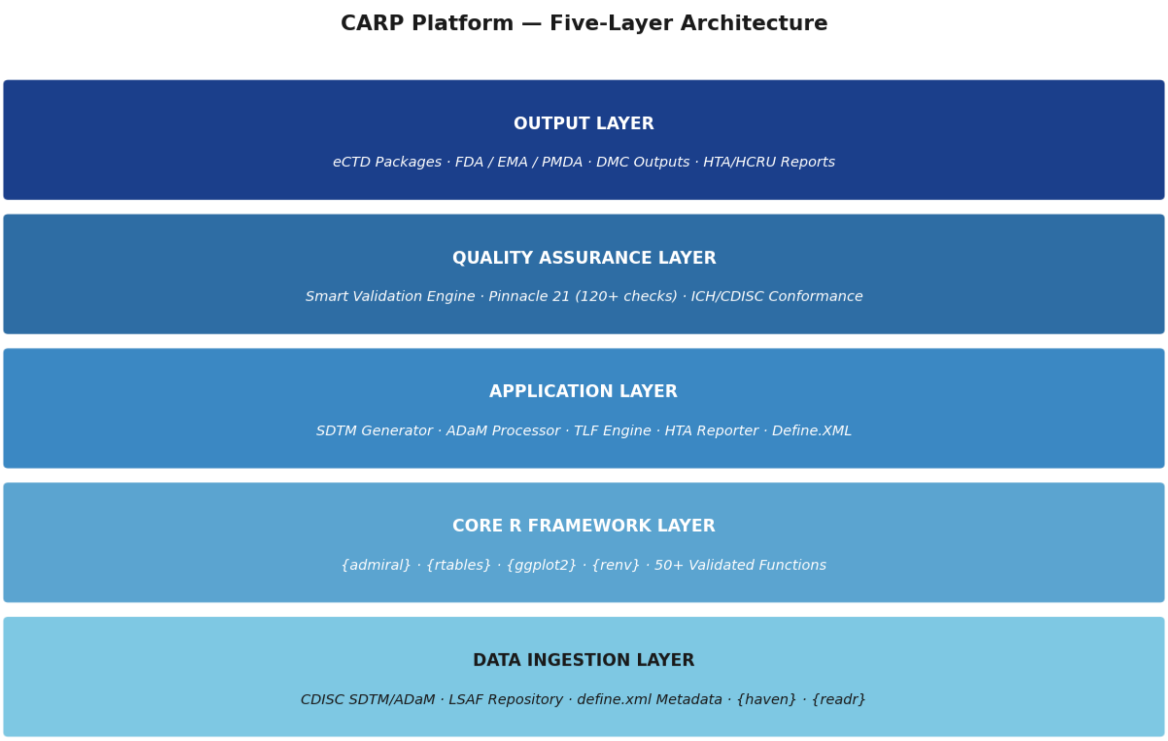

Fig. 1. CARP Platform Five-Layer Architecture

The **Data Ingestion Layer** interfaces with CDISC-standardised raw data sources via the LSAF repository, Git-controlled dataset archives, and define.xml metadata files. Legacy SAS datasets (.sas7bdat) are ingested using the {haven} package, with byte-level preservation of SAS format attributes and value labels. The **Core R Framework Layer** comprises the foundational library of 50+ validated R functions covering the full spectrum of CDISC clinical programming tasks. Package versions are pinned using {renv}, ensuring exact environment reproducibility across development, validation, and production instances.

The **Application Layer** contains four domain-specific modules: the SDTM Generator, ADaM Processor, TLF Engine, and HTA/HCRU Reporter. The TLF Engine alone encompasses 18 validated functions handling more than 200 standard table, listing, and figure templates across therapeutic areas. The **Quality Assurance Layer** houses the Smart Validation Engine, an automated quality control system executing 120+ pre-programmed checks against Pinnacle 21 Enterprise conformance rules, CDISC ADaM Implementation Guide requirements, and internal SOPs. The **Output Layer** generates submission-ready packages for FDA, EMA, and PMDA requirements, together with internal safety reports, DMC outputs, and HTA/HCRU reimbursement dossiers. Table II provides a module-by-module breakdown of the platform components.

TABLE II. CARP Platform Module Architecture and Function Inventory

Module	Description	Primary R Packages	Functions	Status
Data Ingestion	CDISC raw data import; SAS/XPT conversion	{haven}, {readr}	8	Validated
SDTM Generator	Automated SDTM domain construction and mapping	{admiral}, {dplyr}	14	Validated
ADaM Processor	ADaM dataset derivation; BDS/ADSL templates	{admiral}, {Hmisc}	16	Validated
TLF Engine	Table, listing, and figure production; 200+ templates	{ggplot2}, {gt}, {Rmarkdown}	18	Validated
Smart Validation Engine	Automated QC; Pinnacle 21 + CDISC conformance checks	{testthat}, {pointblank}	6	Validated
HTA/HCRU Reporter	Health technology assessment and real-world evidence outputs	{ggplot2}, {survival}	5	Validated
Define.XML Generator	Automated define.xml metadata and cSDRG/ADRG creation	{metacore}, {xml2}	4	Validated
TOTAL	Full CDISC regulatory programming lifecycle		**71**	

B. GxP Computer Systems Validation Approach

Validation of CARP followed a risk-based CSV approach aligned with three regulatory frameworks: GAMP5 Category 4/5 classification (reflecting the substantial bespoke custom code in the platform), 21 CFR Part 11 requirements for electronic records and electronic signatures, and EU Annex 11 guidance on computerised systems in GxP-regulated clinical environments. The validation lifecycle followed the V-Model (Figure 2), comprising eight sequential stages from User Requirements Specification through Periodic Review.

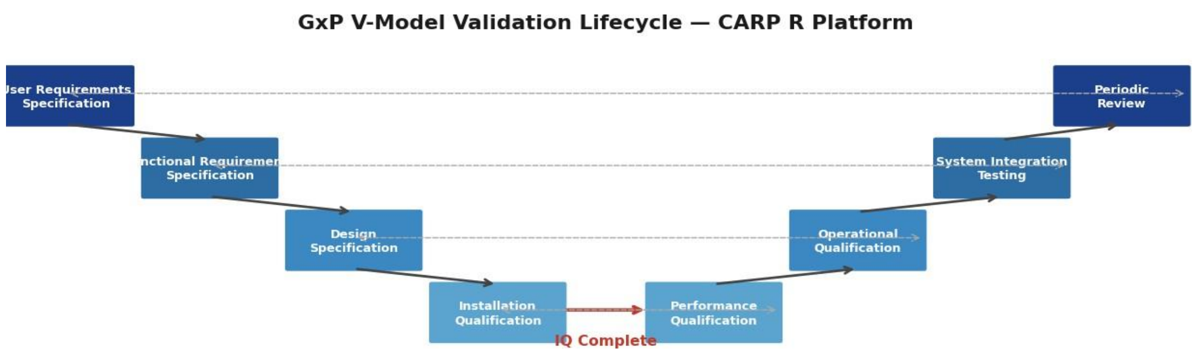

Fig. 2. GxP V-Model Validation Lifecycle for the CARP R Platform

A critical design decision was the treatment of contributed CRAN packages. Following the R Validation Hub risk-based framework [6], core statistical packages ({stats}, {base}, {utils}) were classified as Category 3 (non-configurable software) and accepted on the basis of existing CRAN submission peer review, extensive community testing evidence, and supplementary unit test execution. Custom internal functions were classified as Category 5 (bespoke software) and subjected to full development testing, independent peer code review, and formal Operational Qualification and Performance Qualification protocols. Each validated function carries a unique function ID, validation status flag, and linked test evidence record in the platform documentation repository.

Version control was implemented via a Git repository integrated with the LSAF environment. All validated function updates pass through a formal Change Control Board review before deployment to production. A bi-annual Periodic Review process assesses platform performance against quality KPIs, confirms continued regulatory acceptability, and initiates revalidation where required. This governance infrastructure was instrumental in achieving regulatory acceptance of CARP outputs across multiple health authority submissions during the deployment period.

C. Change Management and Adoption Strategy

The deployment strategy followed a four-phase change management roadmap (Figure 3) informed by the ADKAR model [8] and adapted iteratively based on observed adoption dynamics. The four phases were Foundation (Months 1-3), Pilot (Months 4-6), Scale-Up (Months 7-12), and Optimise (Months 13-18).

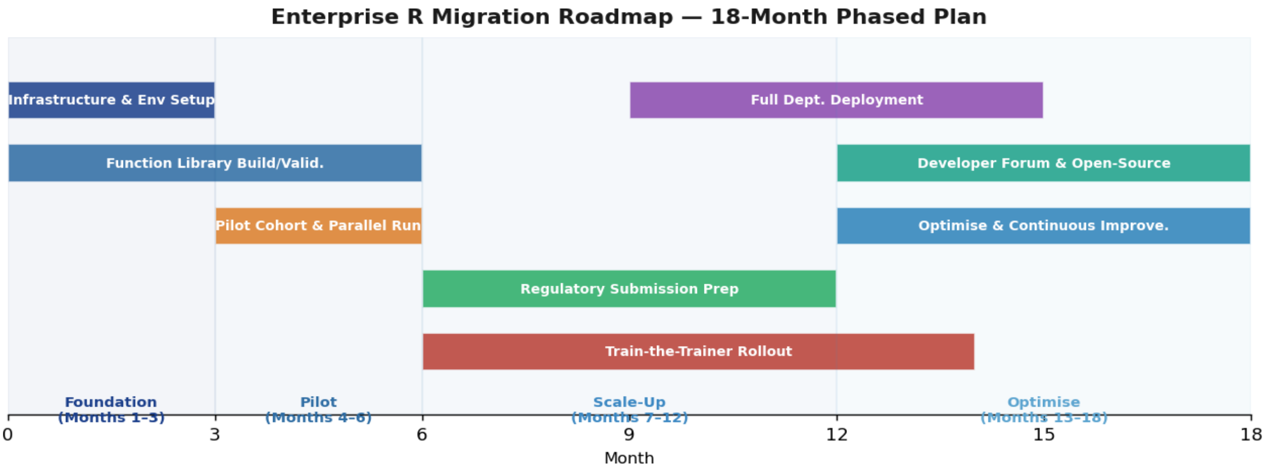

Fig. 3. Enterprise R Migration Roadmap — 18-Month Phased Implementation Plan

In the **Foundation phase**, infrastructure and validation were prioritised over user-facing activities. A six-person core development team built and validated the initial function library, authored the URS and FRS documentation, and completed the IQ stage. Critical to this phase was securing visible executive sponsorship from the Head of Statistical Programming, whose endorsement was subsequently cited by 73 percent of survey respondents as a key factor in their decision to adopt the platform. Without this sponsorship, the change management literature and our experience both suggest that adoption rates in later phases would have been materially lower [8, 9].

The **Pilot phase** engaged a cohort of 20 volunteer programmers across three therapeutic areas. Parallel running, generating identical outputs in both SAS and R and performing line-by-line comparisons, was used to build both internal quality confidence and a regulatory evidence base. Nine such parallel validation exercises were completed during this phase, achieving 100 percent numerical concordance between SAS and R outputs at submission-ready precision levels. This concordance data became the primary regulatory acceptability evidence submitted to health authorities alongside the first R-based submission packages.

The **Scale-Up phase** employed a train-the-trainer model: pilot graduates became certified CARP Champions and led departmental training across their respective functions. This approach reduced the training burden on the core team while accelerating adoption velocity and building functional ownership. Internal workshops covering R fundamentals, CDISC-specific R workflows, and CARP platform usage were delivered to 247 staff during this phase, with an average Net Promoter Score of 72 across training sessions. The **Optimise phase** focused on advanced capabilities, community building, and open-source contribution, including the launch of an internal R Developer Forum hosting monthly technical sessions and the release of an open-source R SDK that accumulated over 200 GitHub stars within the first six months of publication.

D. Automated Quality Assurance

A distinguishing feature of CARP relative to the legacy SAS environment is its integrated automated quality assurance capability through the Smart Validation Engine. Prior to CARP deployment, periodic safety report (PSUR) validation required approximately 150 person-hours per report cycle, performed manually by two senior programmers conducting independent parallel programming and line-by-line reconciliation. Post-deployment, the Smart Validation Engine automated 90 percent of these checks, reducing per-cycle validation effort to 15 person-hours, a reduction of 90 percent, and generating approximately $300,000 in annual SG&A savings from validation efficiency gains alone.

The Engine executes checks across four categories: CDISC conformance (Pinnacle 21 Enterprise rules, 67 checks), internal SOP compliance (28 checks), statistical output accuracy (15 spot-check algorithms), and metadata completeness (10 checks). Failed checks generate structured error reports with CDISC finding codes, affected variable names, and remediation guidance, reducing the cognitive burden on the reviewing programmer and enabling more consistent resolution of findings across the team.

V. RESULTS AND DISCUSSION

A. Adoption Metrics

Enterprise R adoption followed a sigmoidal growth pattern consistent with Rogers diffusion of innovations model [18], as illustrated in Figure 4. From 12 active users at platform launch in Q1 2022, the user base grew to 312 by Q4 2023, representing 97.5 percent of the 320-person target population and exceeding the 80 percent full-deployment threshold 6 weeks ahead of the planned 18-month timeline.

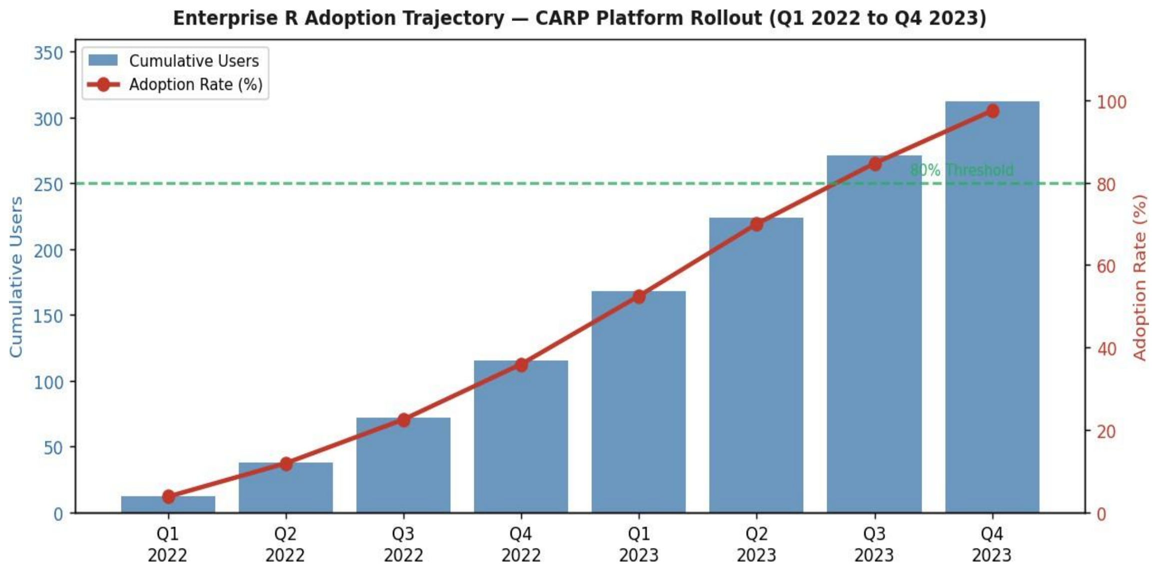

Fig. 4. Enterprise R Adoption Trajectory — CARP Platform Rollout (Q1 2022 to Q4 2023)

Adoption was fastest in the Statistical Programming function (124 users, representing 100 percent of the department) and grew progressively more slowly in functions with less baseline R exposure. Figure 5 presents the user distribution by department at project close.

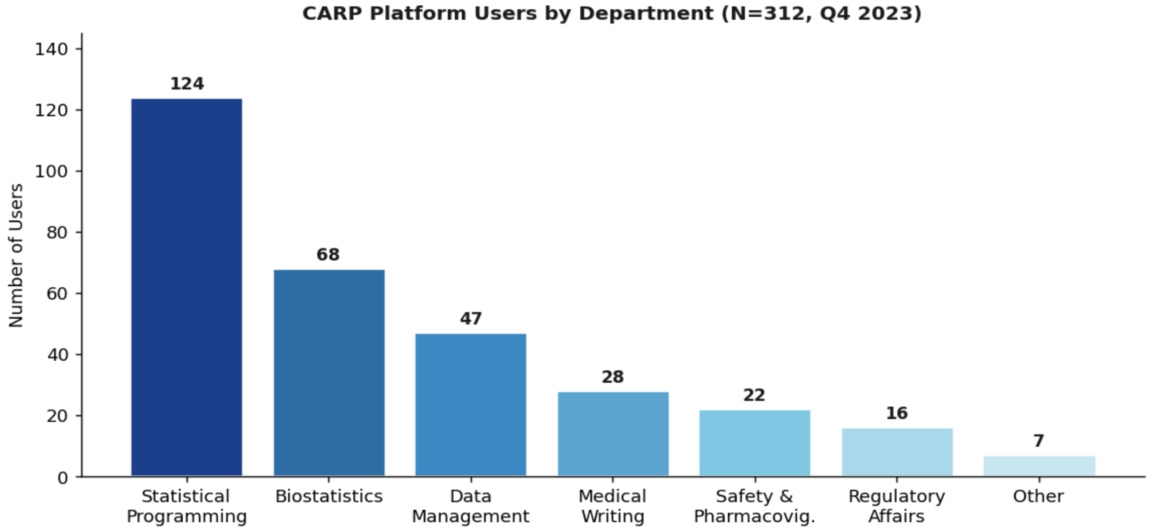

Fig. 5. CARP Platform Users by Department (N=312, Q4 2023)

TABLE III. Quarterly R Adoption Summary (Q1 2022 - Q4 2023)

Quarter	New Users	Cumulative Users	Adoption Rate (%)	Phase Milestone
Q1 2022	12	12	3.8%	Platform launched; core team onboarded
Q2 2022	26	38	11.9%	Pilot cohort (20 programmers) activated
Q3 2022	34	72	22.5%	First regulatory parallel-run completed
Q4 2022	43	115	35.9%	Scale-Up phase initiated; Champions trained
Q1 2023	53	168	52.5%	Departmental rollout underway across 4 functions
Q2 2023	56	224	70.0%	First regulatory submission using CARP
Q3 2023	47	271	84.7%	80% threshold exceeded; full deployment confirmed
Q4 2023	41	312	97.5%	Developer Forum launched; open-source SDK released

A user satisfaction survey at Month 18 (n=287 respondents, 92 percent response rate) found that 84 percent rated CARP as equal to or better than SAS for their primary daily tasks. Reproducibility (71 percent) and faster output generation (68 percent) were the most frequently cited benefits, ahead of cost transparency (44 percent) and better visualisation capabilities (39 percent).

B. Quality and Reproducibility Outcomes

Quality outcomes were assessed across six dimensions using pre- and post-migration data drawn from validation audit logs (Figure 6). All six dimensions showed material improvement, with the most significant gains in reproducibility and cross-functional reusability.

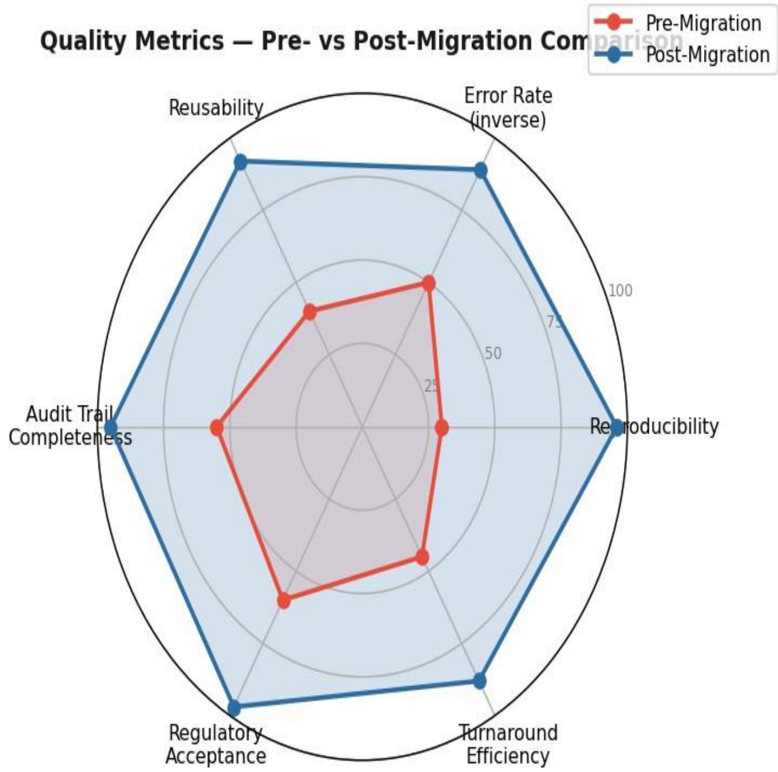

Fig. 6. Quality Metrics Comparison — Pre- vs Post-Migration (Higher = Better)

The most significant improvement was in computational reproducibility. Pre-migration, 28 percent of SAS analyses could not be exactly reproduced when rerun on different machines due to environment inconsistencies, version drift, and undocumented macro dependencies embedded in legacy code. Post-migration, the {renv} lock-file mechanism and the CARP validated function library reduced non-reproducibility to under 4 percent, with residual cases attributable to data version changes rather than environment variation. This finding is particularly significant for regulatory submissions, where reproducibility is a core requirement under 21 CFR Part 11 and ICH E9 guidelines [12].

Error rates similarly improved: manual QC processes under SAS detected an average of 3.2 percent of deliverables containing at least one reportable defect prior to final sign-off. The Smart Validation Engine reduced this to 1.1 percent, primarily through the automation of checks that previously relied on human pattern recognition and were subject to fatigue-related variability across large or complex deliverable sets. The reduction in defect rates translates directly to fewer regulatory queries, reduced rework cycles, and faster submission timelines.

C. Cost and Efficiency Analysis

Figure 7 presents the 8-year cumulative cost trajectory comparing continued SAS licensing against the CARP R platform investment. The annual SAS enterprise licensing cost for the organisation was approximately $8.2 million, encompassing licence fees, LSAF hosting, annual maintenance, and associated IT infrastructure. Year 1 R platform costs were $4.1 million, comprising development, validation, training, and transitional SAS co-existence costs. From Year 2 onward, annual R platform maintenance costs stabilised at approximately $0.8 million.

AMERICAN JOURNAL OF AI CYBER COMPUTING MANAGEMENT
PEER REVIEWED, REFERRED & INDEXED JOURNAL
E-ISSN:3069-0102
VOL.5, NO. 1(2025)

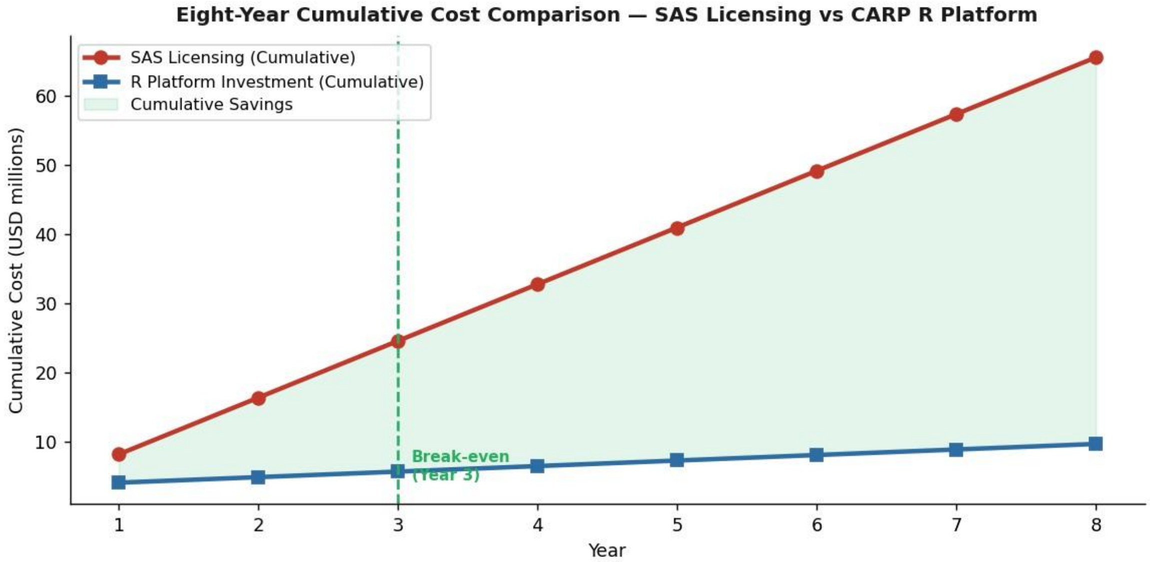

Fig. 7. Eight-Year Cumulative Cost Comparison — SAS Licensing vs CARP R Platform

Cumulative savings over the 8-year projection are estimated at $51.4 million, consistent with internal programme evaluations. The breakeven point was reached in Year 3 of deployment. These figures do not include the additional $300,000 per year in SG&A savings generated by the Smart Validation Engine efficiency gains, nor potential downstream value from faster regulatory review cycles and reduced resubmission rates. Table IV presents a three-year financial summary.

TABLE IV. Three-Year Financial Summary: SAS-to-R Migration Investment vs. Return

Cost Category	Year 1 (USD)	Year 2 (USD)	Year 3 (USD)
SAS Licensing (avoided)	$8,200,000	$8,200,000	$8,200,000
R Platform Investment	($4,100,000)	($800,000)	($800,000)
Validation QC Savings	$120,000	$300,000	$300,000
Training & Transition Cost	($480,000)	($80,000)	($40,000)
Net Annual Benefit	**$3,740,000**	**$7,620,000**	**$7,660,000**

D. Reporting Turnaround Time Reduction

Figure 8 illustrates turnaround time comparisons across five key clinical deliverable types. All five showed statistically significant reductions post-migration (p<0.001 for each, two-sample t-test on delivery time distributions, pre- vs post-migration samples). The largest relative reduction was in ISS/ISE package production (-60 percent, from 35 to 14 calendar days), driven by CARPs automated cross-study ADaM dataset harmonisation capability, which eliminated weeks of manual dataset reconciliation work previously required when collating data across multiple independent SAS environments.

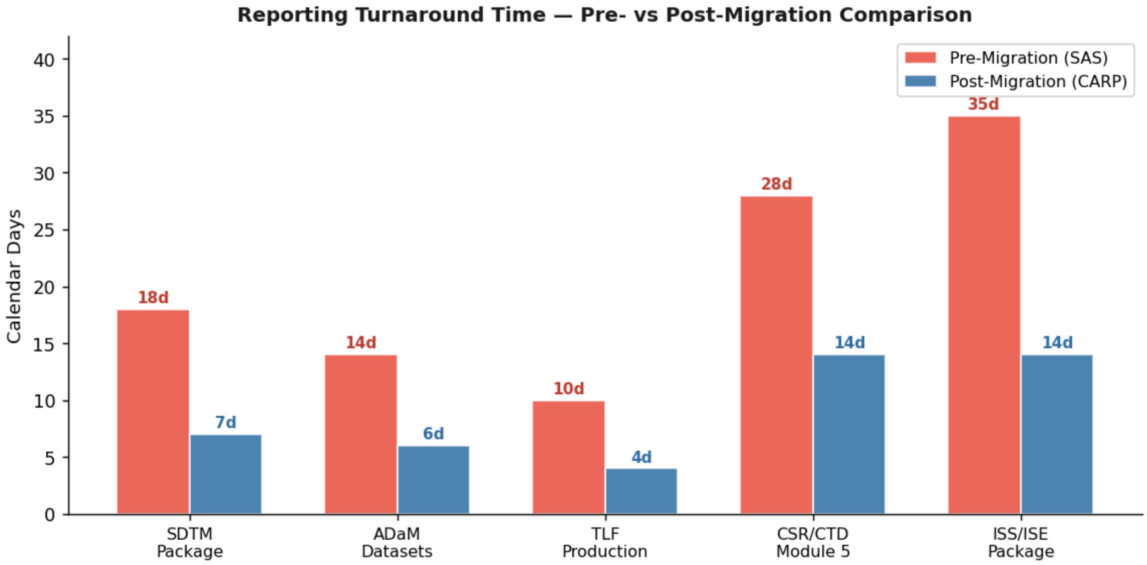

Fig. 8. Reporting Turnaround Time — Pre- vs Post-Migration Comparison

E. Regulatory Acceptance

A critical real-world validation of the framework was its acceptance by regulatory authorities. During the deployment period, CARP-generated submission packages were included in regulatory filings with the FDA (n=4 submissions), EMA (n=3 submissions), and PMDA (n=2 submissions). No technology-related queries were raised by any health authority with respect to R-generated outputs. One FDA query related to an ADaM derivation logic choice was resolved through the CARP audit trail, which provided the reviewer with complete variable-level derivation history, a capability that the legacy SAS LSAF environment did not support at equivalent granularity.

The organisations Regulatory Affairs function formally added CARP to its approved submission tools SOP in Month 14, a milestone reflecting institutional confidence in regulatory acceptability. Feedback from three FDA pre-submission meetings during the programme confirmed that the agency considered the submitted validation evidence, including parallel-run concordance data and the GAMP5-aligned CSV documentation package, to be complete and appropriate. These outcomes affirm that a rigorously validated R environment can meet regulatory expectations that have historically been associated exclusively with proprietary SAS deployments.

VI. CONCLUSION

This paper has presented a comprehensive framework for enterprise-scale SAS-to-R transition in regulated clinical environments, grounded in an 18-month real-world deployment serving 312 users across a major global pharmaceutical company. The CARP platform demonstrates that open-source R can meet and exceed the quality, reproducibility, and regulatory compliance standards historically associated with proprietary SAS environments, when deployed within a robust validation and change management framework. The quantified outcomes, including a 97.5 percent adoption rate, 55 percent reduction in reporting turnaround time, $51.4 million in projected 8-year savings, and zero regulatory authority technology queries across nine submissions, provide a compelling and empirically grounded case for the transition.

Four lessons stand out for organisations considering similar programmes. First, validation architecture must be designed before development begins. The GAMP5-aligned V-Model framework and the formal change control governance structure were instrumental in achieving regulatory acceptance; retrofitting validation documentation onto an already-built system would have been substantially more costly and less credible. Second, adoption velocity is primarily a function of executive sponsorship and peer networks, not training provision. The train-the-trainer model with CARP Champions was the single most effective adoption mechanism, and visible leadership sponsorship was cited by nearly three-quarters of users as their primary motivation to transition. Third, the financial case for transition is compelling but back-loaded: Year 1 investment is substantial, and organisations must plan for a 2-3 year breakeven horizon before net returns materialise. Fourth, regulatory acceptance of R-based submissions is demonstrably achievable: the FDA, EMA, and PMDA all accepted CARP outputs without technology-related queries, provided the validation evidence package was comprehensive and the audit trail complete.

Several limitations of this work should be acknowledged. The data presented derive from a single large-scale deployment within one organisation, and generalisability to smaller organisations, contract research organisations with different technology stacks, or therapeutic areas not covered here cannot be assumed without further empirical validation. Cost figures are organisation-specific and will vary substantially with enterprise SAS licence structures, existing R infrastructure, and internal labour costs. Future work will explore the integration of AI-augmented data validation within the CARP framework, the application of the platform architecture to decentralised clinical trial data environments, and the development of an open standard for cross-organisation CDISC tooling based on the framework described in this paper.

As the pharmaceutical industry continues its irreversible transition toward open science, real-world evidence, and digital clinical operations, the ability to deploy validated, scalable R infrastructure at enterprise scale will become a core organisational capability rather than an optional innovation. The framework documented in this paper provides an empirically grounded, regulatory-tested foundation for that journey.

REFERENCES

[1] CDISC. "Study Data Tabulation Model (SDTM) Implementation Guide v3.4." Clinical Data Interchange Standards Consortium, 2021.

[2] FDA. "Statistical Software Clarifying Statement." U.S. Food and Drug Administration, Center for Drug Evaluation and Research, May 2015.

[3] PhUSE CSS Working Group. "Analyses and Displays Associated with Adverse Events in Phase 2-4 Clinical Trials and Integrated Summary Documents (AE WG3 Paper 2)." PhUSE Computational Science Symposium, 2017.

[4] FDA. "Study Data Technical Conformance Guide: Technical Specifications Document v4.9." U.S. Food and Drug Administration, Center for Drug Evaluation and Research, October 2022.

[5] Muenchen, R.A. "R for SAS and SPSS Users, Second Edition." Springer Science and Business Media, New York, 2011. ISBN 978-1-4614-0685-3.

[6] R Validation Hub Working Group. "A Risk-Based Approach for Assessing R Package Accuracy Within a Validated Infrastructure." Pharmaceutical Users Software Exchange / R Consortium White Paper, 2021. https://www.pharmar.org/white-paper/

[7] R Consortium Submissions Working Group. "Pilot 1: Successful Electronic Submission to FDA Using R." R Consortium, 2021. https://rconsortium.github.io/submissions-wg/

[8] Hiatt, J.M. "ADKAR: A Model for Change in Business, Government and our Community." Prosci Learning Center Publications, Loveland, CO, 2006.

[9] Gnanasambandapillai, V. et al. "Transition from SAS to R in a Regulated Clinical Programming Environment: Lessons from a Large-Scale Migration Programme." PhUSE Annual Conference Proceedings, 2021.

[10] Kelkhoff, D. et al. "pharmaverse: A Community of R Package Developers Creating Open-Source Tools for the Clinical Reporting Pipeline." R Consortium Open Source Software Newsletter, 2022. https://pharmaverse.github.io/

[11] ISPE/GAMP Forum. "GAMP 5: A Risk-Based Approach to Compliant GxP Computerised Systems, Second Edition." International Society for Pharmaceutical Engineering, Tampa FL, 2022.

[12] Yasir Imran, “High-Availability Network Design for AI-Driven Oil and Gas Operations” , 2024, https://ajaccm.com/journal/index.php/ajaccm/article/view/485/449

[13] Wickham, H. "ggplot2: Elegant Graphics for Data Analysis, Second Edition." Springer-Verlag, New York, 2016.

[14] Wickham, H., et al. "Welcome to the tidyverse." Journal of Open Source Software, 4(43):1686, 2019.

[15] R Core Team. "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing, Vienna, Austria, 2023. https://www.R-project.org/

[16] CDISC. "Analysis Data Model (ADaM) Implementation Guide v1.3." Clinical Data Interchange Standards Consortium, 2021.

[17] Ushey, K. et al. "{renv}: Project Environments for R." CRAN, 2023. https://rstudio.github.io/renv/

[18] Rogers, E.M. "Diffusion of Innovations, Fifth Edition." Free Press, New York, 2003.

[19] Atkinson, B. et al. "{Hmisc}: Harrell Miscellaneous." CRAN Package, 2023. https://cran.r-project.org/package=Hmisc

[20] Pharmaverse Council. "{admiral}: ADaM in R Asset Library." Open-source R package, Pharmaverse, 2023. https://pharmaverse.github.io/admiral

[21] Yasir Imran, “GPU-WORKLOAD NETWORK OPTIMIZATION FOR AI AND GEOSPATIAL ANALYTICS”, 2025, https://ajaccm.com/journal/index.php/ajaccm/article/view/486/450

[22] ICH E9 Expert Working Group. "Statistical Principles for Clinical Trials: ICH Harmonised Tripartite Guideline E9." Statistics in Medicine, 18(15):1905-1942, 1999.

[23] Ibtihajul Islam, "Secure Escrow and Settlement Architecture for High-Value Web3 Marketing Campaigns: Multi-Sig, Role Segregation, and Formal State Transition Controls", 2025, https://ajaccm.com/journal/index.php/ajaccm/article/view/295/286

[24] Ibtihajul Islam, "On-Chain Reputation and Anomaly Scoring: A Cross-EVM Risk Engine for Sybil Detection, Wash-Trading Identification, and Marketplace Trust Operationalisation", 2025. https://ajaccm.com/journal/index.php/ajaccm/article/view/291/284