

R and SAS - How They Can Co-Exist? Approaches and considerations in adopting multiple languages The Implementation path for a “best in class” approach depends upon several factors; let’s discuss!

Bill Qubeck, Bill.Qubeck@sycamoreinformatics.com, Rakesh Gotiwale, Rakesh.Gotiwale@instem.com

ABSTRACT

In the pharmaceutical industry, Clinical Study Reporting (CSR) is vital for documenting, analyzing, and communicating trial data. While SAS has been the dominant tool due to its regulatory acceptance and robust capabilities, many organizations are exploring R (and other languages) for its advanced statistical modeling and data visualization capabilities. However, transitioning to R poses challenges, especially for companies heavily reliant on SAS infrastructure. This poster discusses strategies for integrating R into the CSR workflow without disrupting established SAS processes. It explores approaches for both legacy organizations deeply invested in SAS and newer companies with less reliance on it, aiming to enhance data analysis and reporting by leveraging the strengths of both tools efficiently and compliantly.

Type 1: Predominantly SAS Based Organizations

Organizations with a long history of clinical studies tend to have accumulated vast repositories of SAS artifacts, including datasets, programs, analysis, logs, etc.. They typically have well-established global libraries, therapeutic-specific programs, and standard operating procedures (SOPs) aligned with regulatory expectations. Additionally, their staff are well-trained in SAS, with many years of operational experience. Integrating R into such an environment requires careful consideration of how to balance the benefits of R with the deep entrenchment of SAS.

Type 2: Organizations with Minimal SAS Dependency and Legacy Artifacts

In contrast to large, established organizations with entrenched SAS infrastructures, newer organizations often have significantly less dependency on SAS. These organizations may be early in their clinical development lifecycle, with few or no completed clinical studies and therefore have little reliance on established SAS datasets or programs. Without the historical collection of extensive SAS assets, they have the flexibility to adopt modern statistical tools like R without being constrained by legacy systems. However, this flexibility comes with its own set of challenges and considerations when deciding how to integrate SAS and R into their workflows.

Type 1 Organizations: Adoption Considerations for R and Other Languages

Training and Expertise
Adapting to R requires retraining existing staff or hiring R-proficient talent. Upskilling legacy staff is key for a smooth transition. Offering internal training and/or working with external experts can help ease resistance and ensure effective integration.

Migration costs and library maintenance
Fully replacing SAS with R involves significant labor and time, especially when duplicating SAS libraries in R. This leads to higher costs as both libraries must be managed, requiring duplicate efforts for code changes, validation, and maintenance. A more cost-effective approach is to adopt a coexistence model, where SAS is retained for standardized tasks while R is gradually introduced for specific tasks and analyses. This minimizes disruption, reduces duplicative efforts, and allows R to enhance visualization and complex modeling alongside SAS.

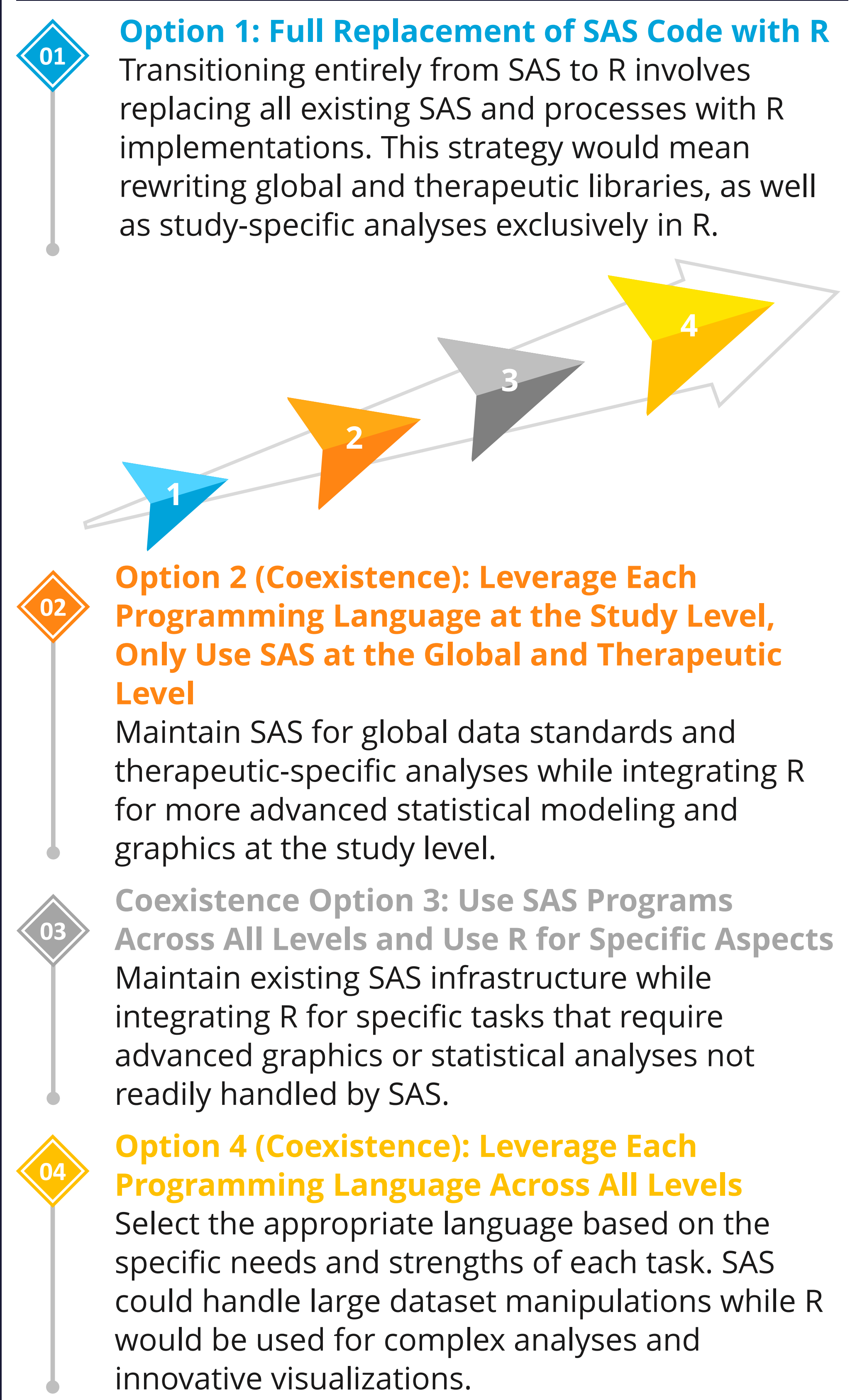
Statistical Computing Environment (SCE)

Must be Capable of Fully Supporting Multiple Languages Out-of-the-Box

It is imperative that the SCE delivers the same capabilities equally well for each programming language in use today and in the future (SAS, R, Python, etc.). It should ensure consistent functionality and support regardless of the language. For example, it must provide:

- Cohesive and Efficient End-User Experience
- Unified Security, Data Access and Sharing
- Results Reproducibility
- Traceability, Lineage for Programs, Data, & Outputs
- Version Control and Dependency Management
- Execution Management and Resource Allocation
- Integrated Development Environment (IDE)
- Workflow Management and Orchestration
- Open-Source Package Management
- Project Team Collaboration

Type 1 Organizations: Four Main Implementation Options



Remember: SAS, R, Python, etc. Can & Do Produce Slightly Different Results...

SAS and R can produce different results in statistical analyses, especially at very high levels of precision (e.g., 0.0001 level). The differences often arise due to:

- **Numerical Precision and Floating-Point Arithmetic:** Variations in how each software handles floating-point calculations can lead to discrepancies.
- **Algorithm Implementations:** SAS and R may use different algorithms or methods for the same statistical procedures.
- **Default Settings and Options:** Differences in default parameter settings (e.g., handling of missing data, convergence criteria) can affect results.
- **Data Handling and Processing:** The way data is imported, processed, and managed can introduce differences.

Starting Point to Incorporate R into Operations

Global Library

Type 1: Not recommended as step 1; there’s significant operational overhead, lack of experience, knowledge, processes, etc.

Type 2: Leveraging ‘fit for purpose’ language approach (R, SAS, & other tools) is “best in class” as step 2 or 3

Therapeutic Library

Type 1: Same as global library; not recommended as step 1

Type 2: Excellent step 2

Study Library

Type 1: Highly recommended as step 2

Type 2: Recommended as step 1; once organization matures implementation of therapeutic & global standards are next steps

Ad hoc & non-GxP

Type 1: Highly recommended as step 1 to incorporate R; lowest risk and costs; leverages ‘fit for purpose’

Type 2: Likely not a major consideration as the focus is on study deliverables (e.g., TFLs for CSRs)

Interoperability and Partner Compatibility

Building internal infrastructure solely around R requires considerations regarding maintaining interoperability with external partners, such as CROs and development partners, that may use SAS. A hybrid approach enables smooth data exchange and flexibility in collaboration. Establishing processes to convert R outputs to SAS formats, while ensuring reproducibility, is essential for effective collaboration and regulatory compliance.

Final Thoughts for Senior Management

Senior management must carefully weigh the trade-offs between flexibility, cost, and regulatory compliance when deciding how to incorporate R into the clinical study reporting process. Each strategy outlined above has its strengths and risks, but the right approach will depend on the organization’s size, regulatory requirements, and long-term goals. By leveraging the differences between SAS and R, companies can create a hybrid model that balances innovation with compliance, driving efficiency and ensuring the highest standards of clinical data integrity.

CONCLUSION

In conclusion, SAS is currently indispensable for its regulatory compliance, standardized processes, and efficient data management, making it the go-to for final clinical study reports and regulatory submissions. Conversely, R shines in its flexibility, cost-effectiveness, and ability to handle advanced statistical techniques, making it a valuable tool for exploratory analysis, modeling, and visualization. A hybrid approach, where both SAS and R are used strategically depending on the task, allows organizations to balance the compliance benefits of SAS with the flexibility and innovation potential of R.