The US Food and Drug Administration (FDA) does not prescribe any particular software for managing or analyzing clinical data, but it does require that all software packages used in clinical studies be fully documented. Their guidance is that “the computer software used for data management and statistical analysis should be reliable and documentation of appropriate software testing procedures should be available[[1]](#footnote-1)”. They encourage researchers to consult with industry experts on the best methods of validating their analytic software but stop short of recommending specific packages.

With this lack of clear guidance, industry leaders are understandably apprehensive of pursuing open-source analytic solutions if it means risking the significant investments they have made investigating new clinical interventions. The risk is lower to just go with SAS for another year, pay the steep bills and endure an antiquated analytic environment. But it doesn’t have to be that way, FDA has not prohibited the use of open-source tools. Companies are just required to include documentation that their analytic environments have been fully validated. Partnering with a third party like ProCogia with expertise in open-source analytics is the most reliable way to make the transition to open-source.

In its simplest definition, a validated environment is simply one where 2 + 2 always equals 4. But in today’s IT environment where operating systems, analytic packages, and their dependencies are updated on different schedules and may include modules that fail to meet any definition of quality or validity. These tools are often necessary in open-source analytics; it does not mean that a particular library or platform is bad, just that it hasn’t met your organizations definition of valid for use in clinical research.

Open-source systems should be evaluated according to a risk-based approach outlined by the R Validation Hub[[2]](#footnote-2). The life cycle of any software package should be completely documented, including testing coverage and integration with your particular system. This approach can be adjusted to meet the risk tolerance of an organization. Many of these metrics, including documentation and testing, can be improved through remediation.

A typical clinical workflow may only require a few individual libraries: dplyr[[3]](#footnote-3), stats, survival[[4]](#footnote-4), and survminer[[5]](#footnote-5). These four packages are widely used and documented. However, these packages include dozens of other R packages to function, each of which may also have additional dependencies. The resulting dependency tree becomes an intimidating mess of vignettes and versioning.

Our approach separates these sourced packages into two groups, those that are *intended for use* and those that are just *imports*.  *Intended for use* packages are those that are vital for the success of the analysis. For example, clinical workflows involving prospective analyses will need careful validation of the survival package, but perhaps not as close inspections of the graphics package on which it depends for Kaplan-Meier curves. This distinction can be murky and should begin with a written assessment of how these packages will be used within the application. Intensive validation can be focused on only those packages and dependencies deemed *intended* f*or use.*

This intensive evaluation will include a thorough assessment of the package. Statistical packages, those that involve modeling or machine learning algorithms, require extra attention because the reproducibility of those results are so important in a clinical environment. Our developers will work closely with the client to meet their existing definitions of quality assurance or work to develop a mutually agreeable definition. Non-statistical packages, including those for general data management, graphics, database systems, and communication, still require adequate testing coverage and definition.

Analytic software is only as accurate as the operating platform on which it depends. Our engineering team supplements each validation with a full review of your system. They will provide you with a comprehensive health check that fully documents all system dependencies, usage statistics, update history, security configuration, and user access. All identified problems can be remediated after consulting with your IT partners.

Once these packages are fully documented and tested, they can be provisioned in a locked repository with limited access to our clients. Our network admins can configure your environments to only include open-source libraries from this validated repository and additional packages can be added to it on demand.

Provisioning a set of open-source tools may seem like a monumental effort requiring help from outside an organization. Partnering with a third party assures an independent validation process supervised by experts in their fields. Our team of consultants at ProCogia is deeply involved in the open-source community, chairing the R Consortium and developing validated solutions for analyzing clinical data within a regulatory environment. We care deeply about open source analytics and the potential benefit these packages can provide the pharmaceutical industry.

Partner with us on your validation project and make the move from old proprietary software to the cutting edge. We can help you make the transition by configuring the environment outlined above. Our team includes experts on code migration to bundle your existing propriety SAS macros into validated and refactored libraries that can serve your organization going forward. We can continue to manage your validated repository and advise you on new development with your data.

1. https://fda.report/media/109552/Statistical-Software-Clarifying-Statement-PDF.pdf [↑](#footnote-ref-1)
2. https://www.pharmar.org/presentations/r\_packages-white\_paper.pdf [↑](#footnote-ref-2)
3. https://cran.r-project.org/web/packages/dplyr/index.html [↑](#footnote-ref-3)
4. https://cran.r-project.org/web/packages/survival/index.html [↑](#footnote-ref-4)
5. https://cran.r-project.org/web/packages/survminer/index.html [↑](#footnote-ref-5)