



Submissions and
Agencies Workstream

Methods and Challenges of R Validation for Regulatory Submission Analysis

Parveen Kumar

www.geninvo.com
North America | UK | Europe | Asia



Definitions

R software	R is a very powerful object-orientated language which can efficiently produce analyses and high-quality graphics. It can be used for high performance computing and allows easy integration with other software. It can also be used for complex analyses when they are not available in SAS.
SDLC	Software Development Life Cycle
Validation	The FDA (2014) defines validation as “Establishing documented evidence which provides a high degree of assurance that a specific process will consistently produce a product meeting its predetermined specifications and quality attributes”.



www.geninvo.com
North America | UK | Europe | India

GENINVO™

Why R?

Why not R

Why R

Development efforts
and challenges with R

Common misconceptions
on regulations

Controlled
workflows/platforms built
for SAS

Interoperability

Access to latest statistical
methodologies

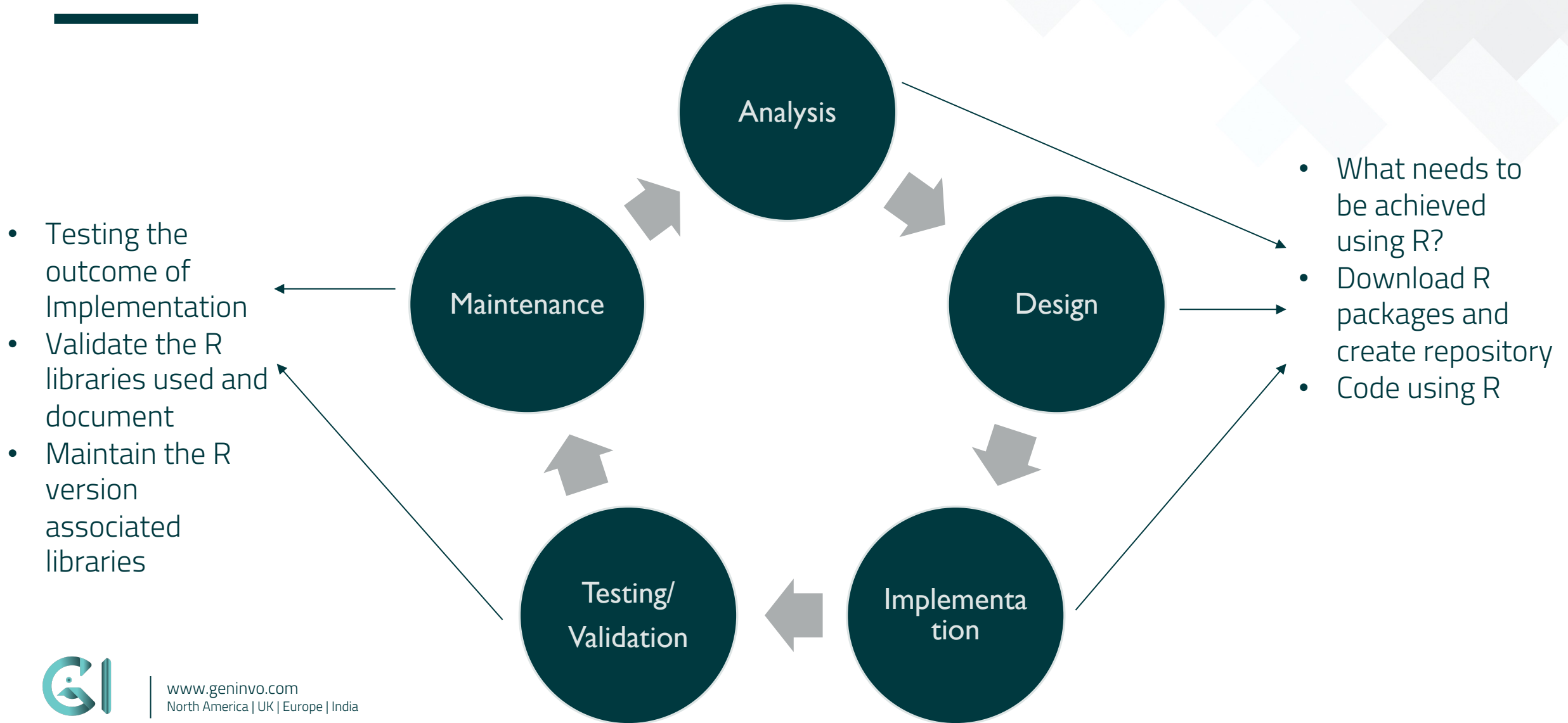
Productivity Gains

Cost

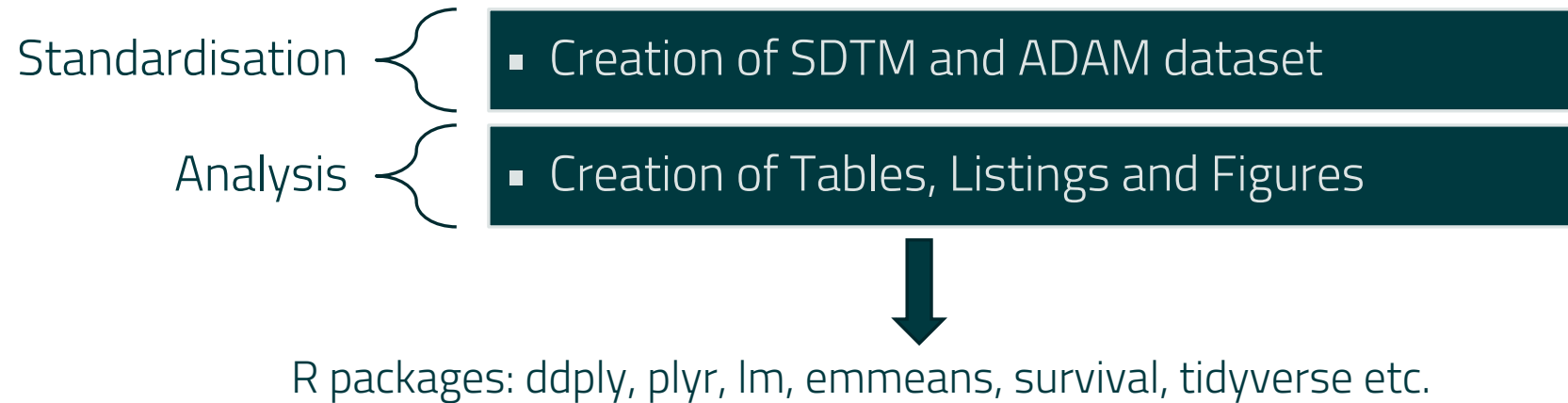


www.geninvo.com
North America | UK | Europe | India

Software Development Life Cycle (SDLC)



Expected Use of R for Clinical Reporting



Validation of R installation

- ▶ Users need to ensure installation validation has been performed according to the company software installation SOPs as applicable.
- ▶ R foundation for Statistical Computing has published “A Guidance Document for the Use of R in Regulated Clinical Trial Environments”
- ▶ R meets all the requirements of 21 CFR Part 11 for validated system if R is used in qualified fashion.
- ▶ Further, R manuals ([CRAN: Manuals \(r-project.org\)](http://CRAN:Manuals(r-project.org))) provides guidance on how to test the installation.
- ▶ R’s source code is managed in a Subversion (SVN) repository with write access limited to the R Core team.



Validation of R packages versus SAS procedures

No statistical software can be purchased as “pre-validated” software in the regulatory context. Validation is entirely on the end user to implement.

SAS procedures are used in industry for many years now. Rather than re-testing, we can make decision to use SAS procedures based on its history of extensive use.

Similarly,

Base R plus Recommended Packages: Many of these R packages have been available for over 20 years and have therefore been open to extensive worldwide peer critique. The R Foundation develops both the base and recommended packages and follows practices that ensures the accuracy of each. These practices include:

- Proper maintenance of the R source code, and control of releases
- Testing the software and identifying issues for the Core Team to address
- The R Core Team hiring highly qualified individuals
- Validation testing each R release against known data and known results, and resolving all errors prior to release



www.geninvo.com
North America | UK | Europe | India

Validation of Contributed R packages

Base R plus Recommended Packages: These package will not be able to fulfill all requirements of analysis.

Contributed Packages:

- R is an open source and packages can be developed by anyone.
- Ensuring the accuracy of contributed packages is necessary
- R CMD provides initial level of validation of packages but that is not enough.

- ▶ SASxport
- ▶ Sas7bdat
- ▶ Dplyr
- ▶ Ggplot
- ▶ Ggsurvplot
- ▶ Survfit
- ▶ lme4



www.geninvo.com
North America | UK | Europe | India

GENINVO™

Risk Assessment (Contributed R packages)

Points to be considered for assessing risk of package

- ▶ Age of package;
 - ▶ How long is it available?
- ▶ Revision history
 - ▶ How many times package is updated and how frequently is it updated?
- ▶ Number of downloads
 - ▶ How many times package is downloaded and any issues reported by users?
- ▶ Reputation of author
 - ▶ How many packages are created by author before?
 - ▶ Is author part of R core team ?
- ▶ Impact of error on analysis
 - ▶ Where will the package be used in the use of R for any submission process?

Validation Strategy (Contributed R packages)

Statistical
Analysis
Packages

- Replicate the similar analysis on SAS for dummy data and compare the results.
- Document the results

Other
packages

- Define the requirements
- Create test scripts
- Perform testing
- Document the results



www.geninvo.com
North America | UK | Europe | India

Maintenance of R packages

- ▶ As R versions and package versions change over time, creating and maintaining R installations becomes more complex.
- ▶ To ensure reproducibility, validated R version and all R packages shall be dockerized with version control.
- ▶ Create a process to document R packages, installation dependencies automatically



Challenges

- ▶ Thorough research is needed to identify R packages needed to support the creation of all type of submission artifacts such as SDTM/ADAM and TLFs. For example: import and export of xpt files, creation of outputs etc.
- ▶ Validation of R packages with complex statistical analysis would need documentation of comparison of results between R and SAS for all possible options. This could be tedious task and very time consuming.
- ▶ Explanation for any differences between SAS and R in analysis results
- ▶ Use of R packages which are updated frequently
- ▶ R packages which are not supported in updated R version



www.geninvo.com
North America | UK | Europe | India

Industry Initiatives and References

- ▶ PharmaR Validation Hub (<https://www.pharmar.org/>)
- ▶ Pharmaceutical companies are setting up internal teams and R repositories (<https://www.rstudio.com/solutions/pharma/>)
- ▶ R core Team (<https://www.r-project.org/doc/R-FDA.pdf>)
- ▶ RStudio: Regulatory Compliance and Validation Issues ([rstudio_compliance_validation.pdf](#))



www.geninvo.com
North America | UK | Europe | India



Ideas Delivered Intuitively

Thank You!

(309) 807-5006 | parveen.k12@geninvo.com | www.geninvo.com

