PhUSE 2014

Paper SD01

R and Revolution R Enterprise in regulatory environments

Eike Nicklas, HMS Analytical Software, Heidelberg, Germany
Andrie de Vries, Revolution Analytics, London, England

ABSTRACT

R, the open source environment for statistics, has been adopted as a data analysis tool by many pharmaceutical companies. Its free license, noncommercial background and fast growth offer many advantages, but also pose challenges to the use in business IT and regulatory environments in particular. Revolution R Enterprise, an R-based platform by Revolution Analytics, offers validated revisions, commercial support and scalable high-speed computations on multiple servers and thus facilitates the use of R in clinical programming and generally in the pharmaceutical industry.
In this contribution, Revolution Analytics and HMS as a service and solution provider introduce R, discuss options for its use in regulatory environments and compare open source R and Revolution R Enterprise in the clinical context. In addition, we discuss tools that support the validation of custom scripts as well as common approaches to quality assurance topics such as unit testing. We illustrate our findings with live demonstrations and concrete examples.

INTRODUCTION

Validation is an essential aspect regarding the employed software in regulatory environments, in particular in the pharmaceutical industry. In order to minimize effort and risk, many companies rely on commercial and well-established software products with high license fees, while the open source environment R emerges as a low cost alternative.

VALIDATION

The FDA summarizes validation as follows: “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.” From this definition, we deduce:

- Not only the employed (software) products need to be validated, but the whole process under discussion. The employed software is only one part of the process. Thus, validation is always necessary, independently of the employed software. However, the effort required for validating the software part can vary and a minimization of this effort is desirable.
- Goal of the validation process is to ensure that the computed results are correct and reproducible.
- The validation result is a document (validation report), that is supposed to convince the regulatory agency (e.g. FDA), such that the above requirements are met.

R

The open source environment R has been increasingly adopted by research divisions of pharmaceutical companies due to its abundance of statistical functions, high-quality plotting capabilities and the large number of “packages” implementing the latest descriptive and predictive analytics algorithms. More than 3 million users worldwide of R have been reported, many of which work at universities and public research institutes. In addition, the usage of R for data science and predictive analytics has grown rapidly in the past years, making it the tool of choice among data miners (see, e.g., Rexer Data Miner Survey). Its open-source nature and noncommercial background offer many advantages such as the lack of license fees, but also pose additional challenges when transferring software from the research departments into the regulatory clinical environment. In recent years, the validation of R has been considered from several perspectives, but still requires effort and is expensive for the adopting companies.

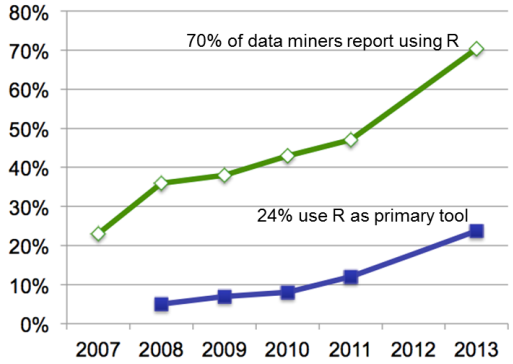

Figure 1: Growth of the usage of R. Source: www.rexeranalytics.com

REVOLUTION R ENTERPRISE

Revolution R Enterprise, an R based platform for enterprise use by Revolution Analytics, aims to ease the utilization of R in business IT by providing e.g. special extensions for big data analytics and better integration into the enterprise IT infrastructure. For the use in regulatory environments, it offers validated revisions, installation qualification protocols, commercial support and allows high speed computations on any data size and in multiple server, cluster and in-database platforms. Thus, it reduces the efforts and risks for pharmaceutical companies, who employ R in clinical programming.

USING R IN A REGULATORY ENVIRONMENT

Many users are still unsure, whether and how R can be used in regulatory environments and what the position of the FDA is regarding the use of R. In the last couple of years, the growing interest in R has led to many new developments by users, R developers and regulatory agencies.
The availability of the following documents and services simplifies the validation process of a software product:

- Quality statement by the vendor / community
- Installation Qualification / Operational Qualification protocols and/or automated tests
- Availability of user support and training

QUALITY STATEMENT BY THE R FOUNDATION

In order to support the use of R in regulatory environments, the R Foundation has released "A Guidance Document for the Use of R in Regulated Clinical Trial Environments" (http://www.r-project.org/doc/R-FDA.pdf) in 2012, which was updated in 2013. This document summarizes the work of the R Foundation for Statistical Computing, introduces R and describes the software development life cycle, change management, testing and release strategies of R. R's Software Development Life Cycle is also summarized in a separate document (http://www.r-project.org/doc/R-SDLC.pdf).

INSTALLATION QUALIFICATION

The installation qualification strategy of R is outlined in the above document. The basic, base and recommended packages, which are supported by the R core development team, include tests, which compare the output of test scripts with a reference output file. These tests can be used to ensure a correct installation of R and also serve as regression tests. The tests can be run using the scripts `testInstalledBasic(scope = "both"), testInstalledPackages("base")` and `testInstalledPackages("recommended")` which are available in the library "tools".

OPERATIONAL QUALIFICATION

The tests used for the installation qualification only compare the string output of test scripts with a reference file. Thus, they ensure that an installation is in agreement with a reference installation, but they do not ensure a logical correctness of the results. For example, if the reference output contains a computation error and the same error is produced by the test script, the tests succeed although the results are incorrect.
For the operational qualification, which is supposed to ensure the correctness of computational results, a different testing strategy is needed, e.g. unit tests. Many R packages already include unit tests, mostly using the unit testing framework RUnit. However, it is required to verify that the tests work correctly and cover all the required functionality. It may be necessary to expand the test coverage by writing your own unit tests.

VALIDATION OF CUSTOM SCRIPTS

The validation process for custom scripts does not depend on the employed technology. However, it can be facilitated by supporting tools, such as integrated development environments (e.g. RStudio and Revolution R Enterprise DevelopR) with debugging capabilities or unit testing (e.g. RUnit). Although the validation of custom scripts is not the focus of this paper, we would like to mention that these tools are available for R.

WHAT DOES THE FDA SAY ABOUT R?

Due to the interest in R for statistical calculations in regulatory environments, the statistical reviewers at the FDA have analyzed the suitability of R for clinical trials. Since 2007 any version of R is allowed to be used on FDA computers by statisticians and is used for product review and research. In his talk at UseR 2007, Mat Soukup (FDA) comes to the conclusions: “Regulations do NOT prevent use of R” and “However, validation, documentation, and accountability are required of sponsors relying on R.”. Since then, R is used regularly at the FDA. Jae Brodsky (FDA) concludes at UseR 2012: “In my experience, R use at the FDA is completely acceptable and has not caused any problems”.

HOW ARE THINGS DIFFERENT IN REVOLUTION R ENTERPRISE?

As discussed in the previous sections, the R Foundation along with the community have made a lot of progress in supporting the use of R in regulatory environments. Revolution R Enterprise is an R based platform for enterprise use developed by Revolution Analytics. It includes the base and recommended packages of open source R, but also provides additional proprietary packages, tools and documents that support the validation of R.

QUALITY STATEMENT BY REVOLUTION ANALYTICS

The development quality assurance process of Revolution Analytics is detailed in a separate document supplied with Revolution R Enterprise. It summarizes, e.g. the configuration and change management, the release process and release criteria, as well as the testing strategy and tools provided for the installation verification process.

INSTALLATION / OPERATIONAL QUALTIFICATION PROTOCOL

Revolution Analytics provides an Installation/Operational Qualification (IQ/OQ) procedure template for Revolution R Enterprise installations, which supports the validation process by serving as a basis for a standard operating procedure. It provides details about the installed packages and references to related documents. Installation instructions are provided both for Windows and Linux along with a detailed installation verification protocol, which includes a step-by-step listing of the necessary manual tests and expected results. Part of this IQ/OQ procedure is a run of the Revolution Analytics Validation Suite.

REVOLUTION ANALYTICS VALIDATION SUITE

Revolution R Enterprise ships with a validation function `RevoIOQ` that can be used to verify a correct installation and operation of all bundled packages. It runs an extended version of the installation tests of R packages mentioned above along with specifically designed unit tests. The result of each run is summarized in a validation report. Additional unit tests can be added on demand.

REPRODUCIBLE R TOOLKIT

While the modular structure of R and it easy extensibility with packages has many advantages, it also poses challenges in a regulatory environment, where a high consistency of the software is required in order to ensure reproducible calculations. This issue is addressed by the Reproducible R Toolkit, which provides mechanisms to install specific versions of R packages available on CRAN and thus ensures that computations can be performed under the same conditions in the future.

CONCLUSION

In conclusion, we discussed some aspect of using R in regulatory environments. In recent years, the R Foundation has made several adjustments that facilitate the validation of R. Revolution R Enterprise further supports the validation process by providing IQ/OQ protocols and an automated validation suite. The new Reproducible R Toolkit allows the execution of R code under controlled and reproducible conditions and is a further step in reducing the effort required for validating R code used in regulatory environments.

ACKNOWLEDGMENTS

The authors acknowledge fruitful discussions with and support from Friedrich Schuster (HMS) and Evgeniya Ashmarina (HMS).

RECOMMENDED READING

The following papers and presentations further discuss validation aspects of open source R:

- Peter Bewerunge (HMS): How much is it? Validation of Open-Source-Software Using the example of R, PhUSE 2010
  http://www.phusewiki.org/docs/2010/2010%20PAPERS/RG03%20Paper.pdf
- Mat Soukup (FDA): Using R: Perspectives of a FDA Statistical RevieweR, UseR 2007
  http://user2007.org/program/presentations/soukup.pdf
- Jae Brodsky (FDA): Some Challenges of Using R in a Regulatory Environment, UseR 2012
  http://blog.revolutionanalytics.com/downloads/FDA-Janice-Brodsky-UseR-2012.pdf
- Ian Francis (Biop): R Validation for Life Sciences
  http://www.baselr.org/presentations/R%20Validation%20in%20Life%20Sciences.ppt

Further information on this topic as well as the offerings by Revolution Analytics regarding validation is found in the Revolutions blog:

- Analyzing clinical trial data for FDA submissions with R
  http://blog.revolutionanalytics.com/2009/01/analyzing-clinical-trial-data-with-r.html
- R, the FDA, and clinical trials
  http://blog.revolutionanalytics.com/2009/02/using-r-in-the-pharmaceutical-industry.html
- Clinical Reporting with R
  http://blog.revolutionanalytics.com/2010/03/clinical-reporting-with-r.html
- FDA: R OK for drug trials
  http://blog.revolutionanalytics.com/2012/06/fda-r-ok.html

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Dr. Eike Nicklas
HMS Analytical Software GmbH
Rohrbacher Straße 26
Heidelberg / 69115
Work Phone: +49 6221 6051176
Fax: +49 6221 6051742
Email: eike.nicklas@analytical-software.de
Web: http://www.analytical-software.de/en/start/

Andrie de Vries
Revolution Analytics Limited
22 Wycombe End
Beaconsfield, Bucks / HP9 1NB
Work Phone: +44 7968 065 151
Email: andrie @revolutionanalytics.com
Web: http://revolutionanalytics.com/