10min

Numerical validation as a critical aspect in bringing R to the Clinical Research

Adrian Olszewski **Olshevski**

Principal Biostatistician at **2KMM**

The R/Pharma 2020 Conference

www.2kmm.eu

aolszewski@2kmm.pl

13th Oct 2020

DISCLAIMER

All trademarks, logos of companies and names of products

used in this document

are the sole property of their respective owners

and are included here for informational, illustrative purposes only,

which falls within the *nominative fair use.*

This presentation is based exclusively on information

publicly available on the Internet under provided hyperlinks.

If you believe your rights are violated, please email me:

aolszewski@2kmm.pl

Agenda

- **Introduction**
  - Who, Why, What?
  - Motivational story
- **The use of R in Clinical Research – myths and facts**
- **Is the situation really that serious?**
- **How to validate R**
  - Obstacles
  - What would help?
- **Summary**

- The 2KMM - a small Polish CRO with global reach, entirely based on R
  - Clinical Research biostatistics: full coverage from **trial design** to **final report**
    - → Real case experience: What I discuss here may likely concern you as well
  - **We do**: observational + RCT (not submitted to the FDA) trials - **21 so far**
  - **We aim at**: RCT (+ CDISC) submitted to the FDA; started
- We use R for:
  - Trial design (classic and adaptive)
  - Data querying, making data sets (own format & CDISC experimental stage)
  - Trial data analysis (full coverage) + validation
  - Producing T/F/Ls and automated report generation (DOCx via **officer + flextable**)
  - Auxiliary, supportive tools and analyses (data review, investigations)

- A very strange situation takes place:
  - Both S[1976/1980] & R[1993/1997] constituted a de facto industry standard in data analysis
  - R is used everywhere, especially in biosciences: epidemiology, medicine, ecology
  - In Pharma R was used for years – **silently**. Recently – R got *reborn officially.*
  - Main areas of use: trial design, PK & PD, simulations, R&D, reporting, graphics

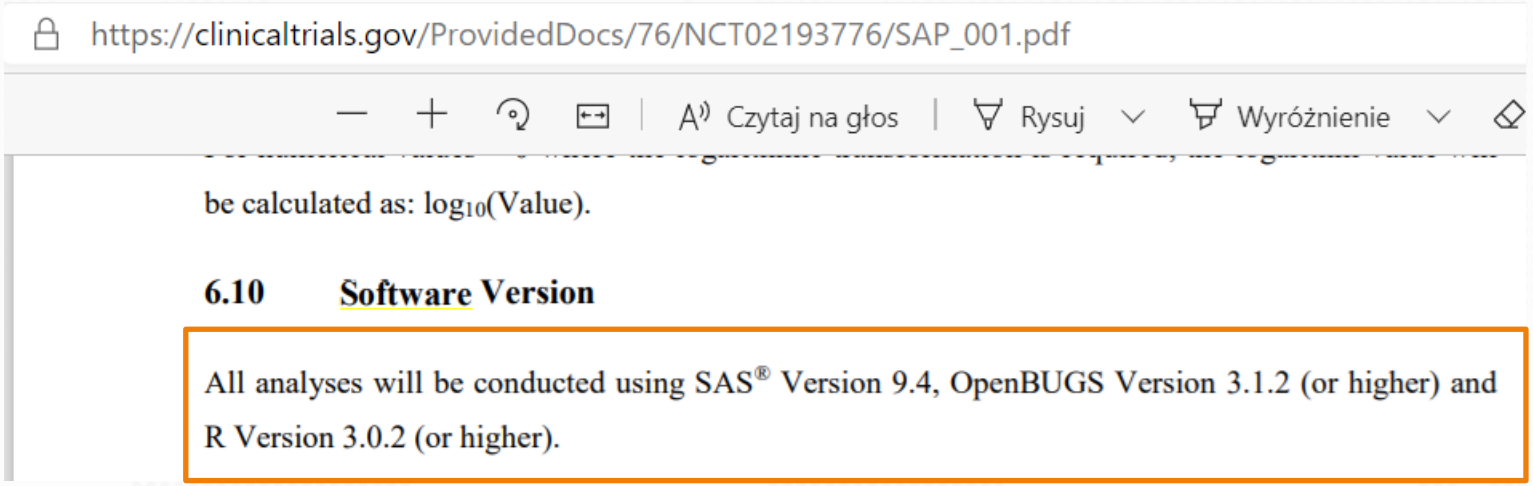
Google: ***site: clinicaltrials.gov AND SAP AND ("r-project" OR "R version“ OR …..)***

SAP:

- https://clinicaltrials.gov/ProvidedDocs/76/NCT02193776/SAP_001.pdf
- https://clinicaltrials.gov/ProvidedDocs/67/NCT01720667/SAP_000.pdf
- https://clinicaltrials.gov/ProvidedDocs/42/NCT02252042/Prot_SAP_000.pdf
- https://clinicaltrials.gov/ProvidedDocs/48/NCT01784848/SAP_001.pdf
- https://clinicaltrials.gov/ProvidedDocs/16/NCT04122716/SAP_000.pdf
- https://clinicaltrials.gov/ProvidedDocs/79/NCT03533179/SAP_000.pdf
- https://clinicaltrials.gov/ct2/show/NCT03797118
- https://clinicaltrials.gov/ProvidedDocs/15/NCT03938415/Prot_SAP_000.pdf
- https://clinicaltrials.gov/ProvidedDocs/48/NCT03135548/SAP_001.pdf
- https://clinicaltrials.gov/ProvidedDocs/79/NCT03098979/SAP_000.pdf
- https://clinicaltrials.gov/ProvidedDocs/65/NCT03702465/SAP_001.pdf

FDA:

- https://www.fda.gov/media/132457/download
- https://www.accessdata.fda.gov/drugsatfda_docs/nda/2009/022129s000_StatR.pdf
- https://www.fda.gov/media/99313/download
- https://www.fda.gov/media/114272/download
- https://www.fda.gov/media/70028/download



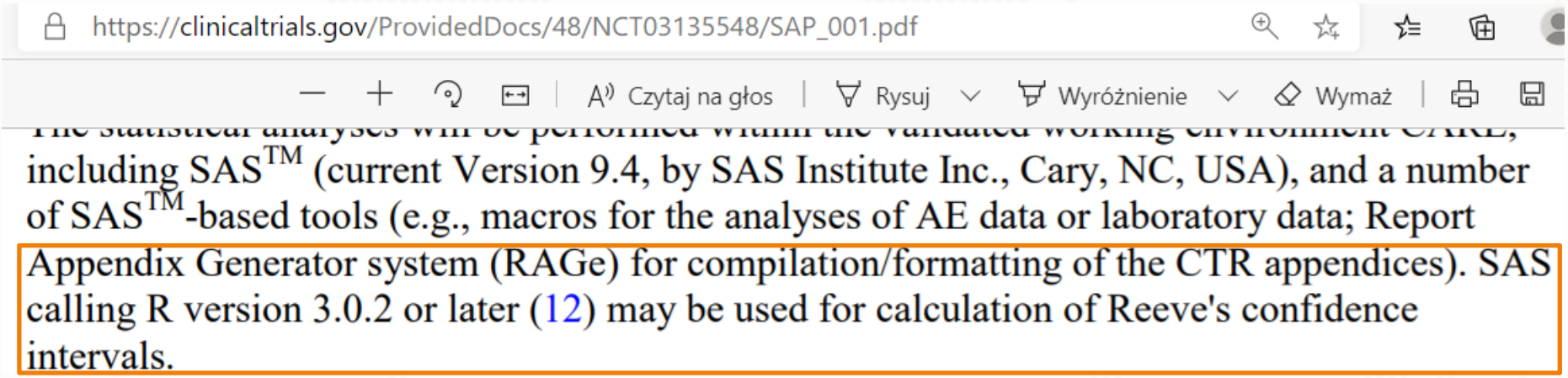
https://clinicaltrials.gov/ProvidedDocs/76/NCT02193776/SAP_001.pdf

Czytaj na głos | Rysuj | Wyróżnienie

be calculated as: $\log_{10}$(Value).

6.10 Software Version

All analyses will be conducted using SAS® Version 9.4, OpenBUGS Version 3.1.2 (or higher) and R Version 3.0.2 (or higher).



https://clinicaltrials.gov/ProvidedDocs/48/NCT03135548/SAP_001.pdf

Czytaj na głos | Rysuj | Wyróżnienie | Wymaż

including SAS™ (current Version 9.4, by SAS Institute Inc., Cary, NC, USA), and a number of SAS™-based tools (e.g., macros for the analyses of AE data or laboratory data; Report Appendix Generator system (RAGe) for compilation/formatting of the CTR appendices). SAS calling R version 3.0.2 or later (12) may be used for calculation of Reeve's confidence intervals.

This Document is incorporated by reference into the following Guidance Document(s):

Study Data Technical Conformance Guide

For questions regarding this technical specifications document, contact CDER at cder-edata@fda.hhs.gov or CBER at cber.cdisc@fda.hhs.gov

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

November 2019

Contains Nonbinding Recommendations

U.S. FOOD & DRUG ADMINISTRATION

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

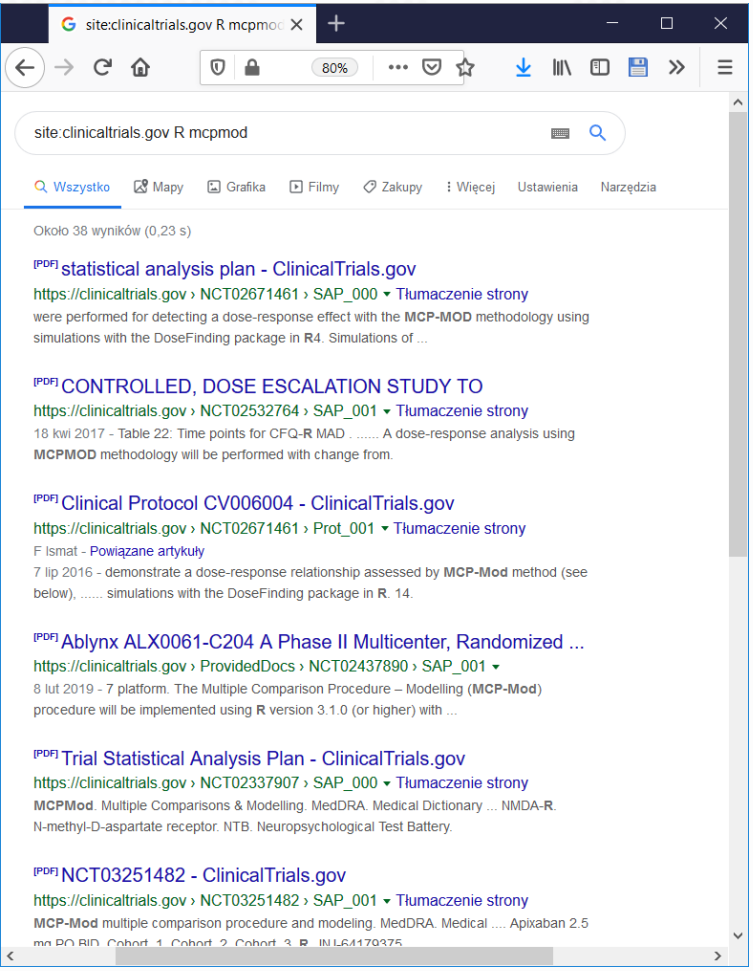
CLINICAL STUDIES

NDA/BLA #:	sNDA 207986
Supplement #:	Supplement-2
Drug Name:	OTIPRIO (ciprofloxacin otic suspension)

[5] Technical note: The sponsor computed unconditional CIs using the *exact riskdiff* statement in SAS proc freq. The reviewer computed unconditional CIs using the function *uncondExact2x2* (with arguments *method = "simple", tsmethod="central"*) from the R package *exact2x2*. Both software gave very similar results, with the R function's CIs being contained within the SAS CIs. For a general discussion of unconditional CIs, see Agresti, A. (2013). *Categorical Data Analysis*. 3rd ed., page 609.

Table 2: Code for Creating ts.xpt Using R : Option B - Using the SASxport Package

R Package	Clinical Study	Non-clinical Study
Option B: Using the SASxport Package	##Load package## library(SASxport) library(Hmisc) ##Create data file## abc<-data.frame(STUDYID="XYZ123", TSPARMCD="SSTDTC",	##Load package## library(SASxport) library(Hmisc) ##Create data file## abc<-data.frame(STUDYID="XYZ123", TSPARMCD="STSTDTC",



OFFICE OF CLINICAL PHARMACOLOGY
DIVISION OF PHARMACOMETRICS

Application	Request for Qualification of MCP-Mod as an efficient statistical methodology for model-based design and analysis of Phase II dose finding studies under model uncertainty
Applicant	Janssen Pharmaceuticals and Novartis Pharmaceuticals
Application date	22 April, 2015
OCP Division	Division of Pharmacometrics
OCP Reviewer	Dinko Rekić, MSc(Pharm), Ph.D.
Concurring reviewers	Yaning Wang, Ph.D. Deputy Director, Division of Pharmacometrics Vikram Sinha, Ph.D. Director, Division of Pharmacometrics

OCP: Office of Clinical Pharmacology

'Fit for Purpose': FDA recognises MCP-Mod's utility to improve dose finding

Methodology incorporated into ICON's ADDPLAN® DF platform for dose finding

MCP-Mod, a powerful statistical tool for reliably predicting optimal dose ranges of new drugs for future confirmatory trials, has been deemed "fit for purpose" by the U.S. FDA. The tool could reduce costly Phase III failures and post-approval dose adjustments.

Social Sharing

gsDesign Explorer to Optimize Merck's Clinical Trial Process

By CIOReview | Monday, April 7, 2014

Tweet Share Share

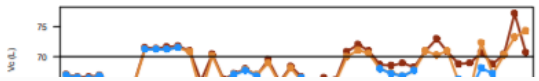
FREMONT, CA: Today, most of the pharmaceutical firms face hurdles in the clinical trial drug development process. They often waste money and time by tirelessly analyzing massive amounts of mission critical data. Aimed at dealing with these kinds of obstacles, Merck, a pharmaceutical firm, has started implementing Revolution Analytics' gsDesign Explorer graphical user interface (GUI).

nlmixr: an open-source package for pharmacometric modelling in R

Rik Schoemaker[1], Yuan Xiong[2], Justin Wilkins[1], Christian Laveille[3], Wenping Wang[4]
[1]Occams, The Netherlands, [2]Certara Strategic Consulting, USA, [3]Calvagone, France, [4]Novartis Pharmaceuticals, USA

Now on github!
https://github.com/nlmixrdevelopment/nlmixr

Aims



Results

"We use R for adaptive designs frequently because it's the fastest tool to explore designs that interest us. Off-the-shelf software, gives you off-the-shelf options. Those are a good first order approximation, but if you really want to nail down a design, R is going to be the fastest way to do that."

Keaven Anderson
Executive Director, Late Stage Biostatistics
Merck

Publicly available sources:

https://pharma-life-sciences.cioreview.com/news/gsdesign-explorer-to-optimize-merck-s-clinical-trial-process-nid-1305-cid-36.html

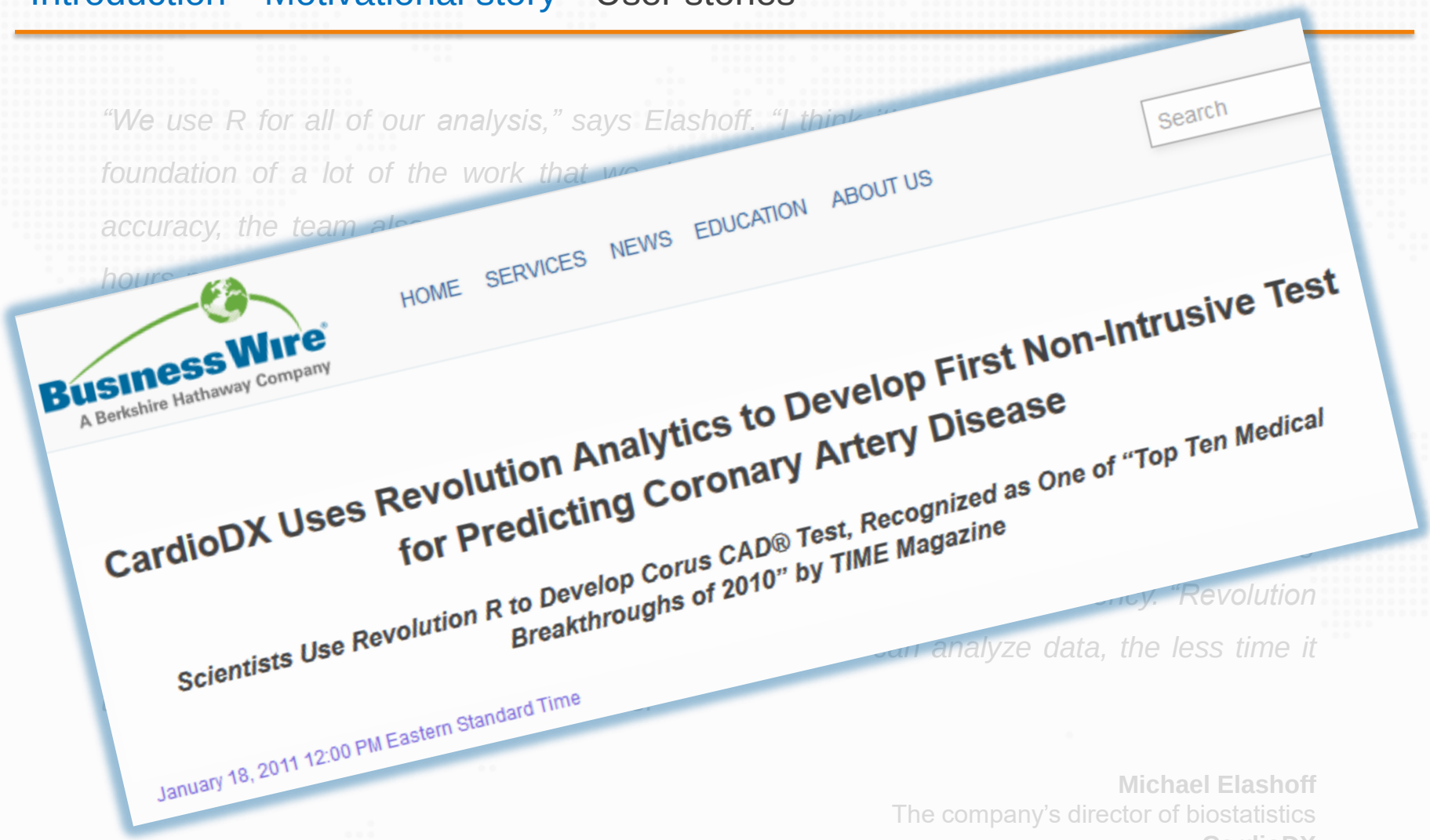
Google Books: Big Data for Big Pharma: An Accelerator for The Research and Development Engine?

"De facto, R is already a significant component of Pfizer core technology. Access to a supported version of R will allow us to keep pace with the growing use of R in the organization, and provides a path forward to use of R in regulated applications."

James A. Rogers Ph.D.
Associate Director, Nonclinical Statistics Group
Pfizer

Publicly available sources:

https://www.featuredcustomers.com/vendor/revolution-analytics-1/customers/pfizer



"We use R for all of our analysis," says Elashoff. "I think ...
foundation of a lot of the work that we ...
accuracy, the team also ...
hours ...
... analyze data, the less time it ...

Michael Elashoff
The company's director of biostatistics
CardioDX

Publicly available sources:

https://www.businesswire.com/news/home/20110118006656/en/CardioDX-Revolution-Analytics-Develop-Non-Intrusive-Test-Predicting

"We use R for all of our analysis," says Elashoff. "I think it's fair to say that R really is the foundation of a lot of the work that we do." To speed up the process without sacrificing accuracy, the team also uses ***Revolution R*** *analytic products. "We use R seven or eight hours per day, so any improvement in speed is helpful, particularly when you're looking at a million biomarkers and wondering if you'll need to re-run a million analyses."*

Open-source R packages enable the biostatisticians at CardioDX to run a broad range of analyses, accurately and effectively, on a routine basis. Adding Revolution R products to the mix improves processing speeds and makes it easier to crunch large data sets. Accelerating the analytic process reduces ov erall project time, increasing the team's efficiency. "Revolution R is faster than regular R," says Elashoff. "The faster we can analyze data, the less time it takes us to build our diagnostic algorithms."

Michael Elashoff
The company's director of biostatistics
CardioDX

Publicly available sources:
https://www.featuredcustomers.com/media/CustomerCaseStudy.document/revolution-analytics-1_cardiodx_8284.pdf

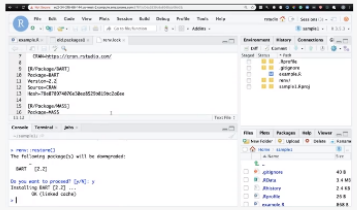
https://rstudio.com/solutions/pharma/

janssen

Janssen Pharmaceuticals is using R and RStudio to address high-performance analytical needs, prioritizing data science workflows that ensure reproducibility and FDA compliance.

"We are also using the enterprise version of RStudio. In case there is an FDA audit and we need to reproduce analysis, we can easily containerize simulations, deploy them internally, forget about them, and come back to them when necessary."

- Satish Murthy, IT Manager Janssen R&D

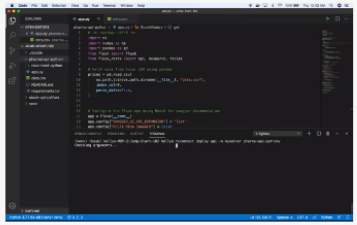


Roche Genentech

Roche & Genentech use RStudio to enable collaboration between hundreds of quantitative scientists at every stage of the drug discovery lifecycle.

"Adapting to the rapidly changing requirements of science requires collaborative software development across the enterprise, industry and field."

- Michael Lawrence, Scientist, Genentech Research and Early Development



NOVARTIS

Novartis uses R to effectively communicate data analysis with stakeholders across the organization.

"Effective use of visualizations enables clear and impactful communication, elevates our influence without stakeholders, and facilitates informed decision making."

- A very strange situation takes place:
  - Both S[1976/1980] & R[1993/1997] constituted a de facto industry standard in data analysis
  - R is used everywhere, especially in biosciences: epidemiology, medicine, ecology
  - In Pharma R was used for years – **silently**. Recently – R got *reborn officially.*
  - Main areas of use: trial design, PK & PD, simulations, R&D, reporting, graphics.

- Then what's wrong if it's so good?
  - Many praise R as the right choice for advanced data analysis
  - They rely on R in trial design (**if failed, entire trial may fail too**) or PK (**toxicity**!)
  - R is used in research and development, decisions are made based on the results
  - **But when it comes to run t.test() for a submission – everyone hesitate**

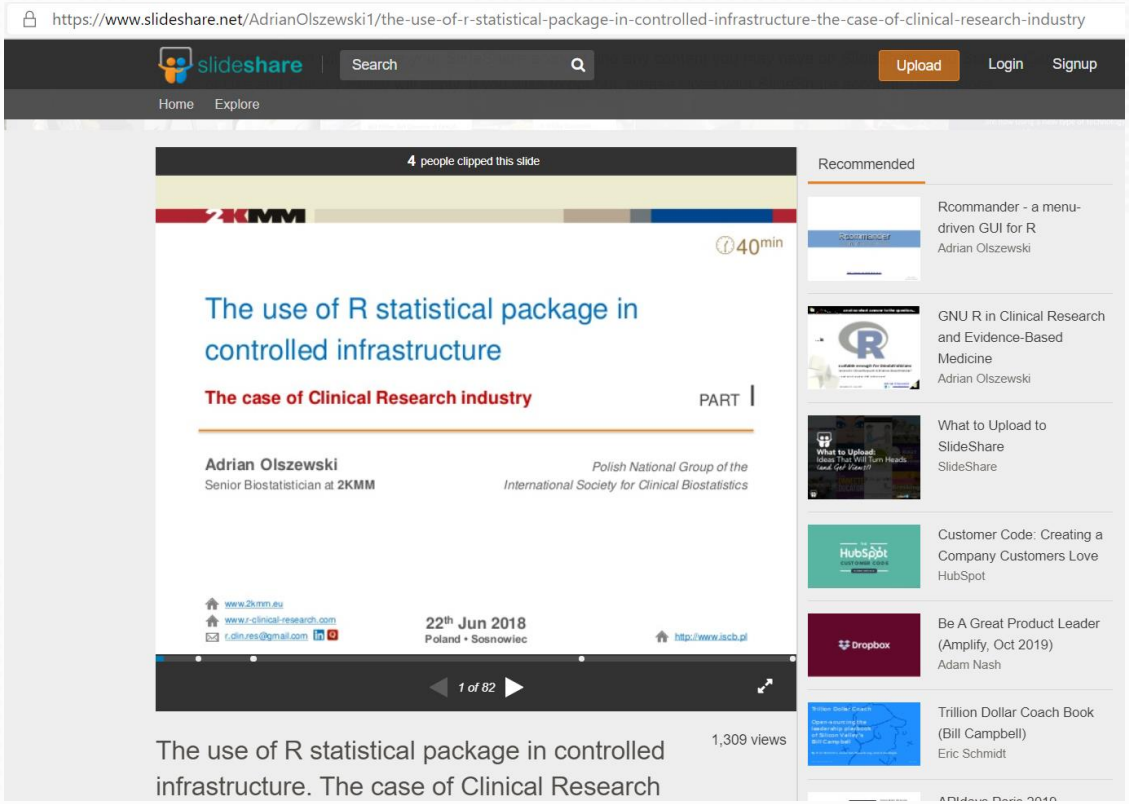
Are they right?

Who is right and…

…is it possible to use R in controlled environment?

https://www.slideshare.net/AdrianOlszewski1/the-use-of-r-statistical-package-in-controlled-infrastructure-the-case-of-clinical-research-industry

Please find the linked presentation for more detailed list of myths and facts



https://www.slideshare.net/AdrianOlszewski1/the-use-of-r-statistical-package-in-controlled-infrastructure-the-case-of-clinical-research-industry

- Quick summary of the presentation
  - Yes, R can be used in Clinical Research, **including submissions**
  - Yes, FDA has nothing against that
  - R (like ANY other software) has to be **validated and properly documented**

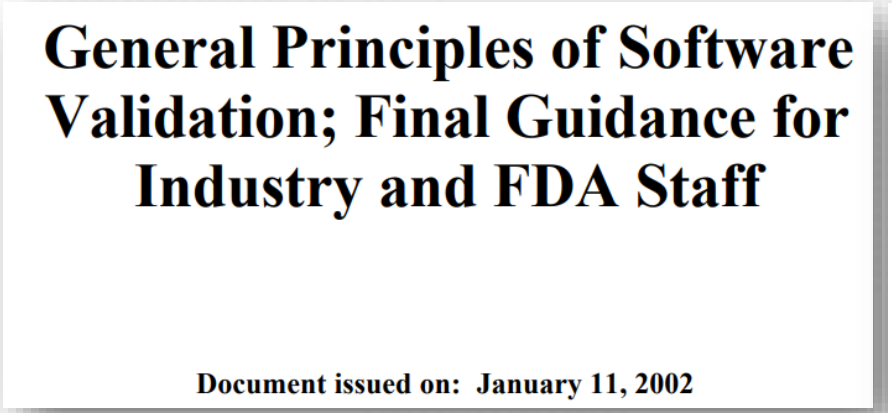
Statistical Software Clarifying Statement

FDA does not require use of any specific software for statistical analyses, and statistical software is not explicitly discussed in Title 21 of the Code of Federal Regulations [e.g., in 21CFR part 11]. However, the software package(s) used for statistical analyses should be fully documented in the submission, including version and build identification.

As noted in the FDA guidance, *E9 Statistical Principles for Clinical Trials* (available at http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm), "The computer software used for data management and statistical analysis should be reliable, and documentation of appropriate software testing procedures should be available." Sponsors are encouraged to consult with FDA review teams and especially with FDA statisticians regarding the choice and suitability of statistical software packages at an early stage in the product development process.

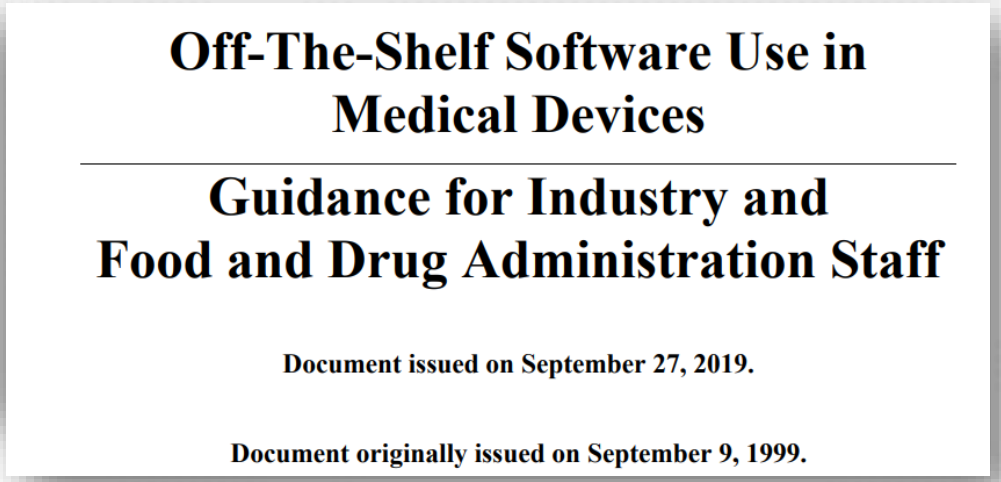
https://www.fda.gov/downloads/forindustry/datastandards/studydatastandards/ucm587506.pdf

The process of validation of the software


General Principles of Software Validation; Final Guidance for Industry and FDA Staff

Document issued on: January 11, 2002

https://www.fda.gov/media/73141/download


Off-The-Shelf Software Use in Medical Devices

Guidance for Industry and Food and Drug Administration Staff

Document issued on September 27, 2019.

Document originally issued on September 9, 1999.

https://www.fda.gov/media/71794/download

[…] FDA considers software validation to be: "confirmation by examination and provision of objective evidence that software specifications conform to user needs and intended uses, and that the particular requirements implemented through software can be consistently fulfilled."

2.1. APPLICABILITY

This document […] can be applied ***to any software*****.**

[…]

This document ***does not specifically identify which software is or is not regulated***

2.4. QUALITY SYSTEM REGULATION VS PRE-MARKET SUBMISSIONS

[…]

The management and control of the software validation process ***should not be confused with any other validation requirements****, such as process validation for an automated manufacturing process*

3.1.1 Requirements and Specifications

[…]

design input ***requirements must be documented****, and that specified requirements must be verified*

[…]

Success in accurately and completely documenting software requirements is a crucial factor in successful validation of the resulting software.

*A **specification** is defined as "a **document that states requirements**."*

3.1.1 Requirements and Specifications

[…]

*There are many different kinds of written specifications, e.g., **system requirements** specification, **software requirements** specification, **software design** specification, **software test** specification, **software integration** specification, etc*

3.1.2 Verification and Validation

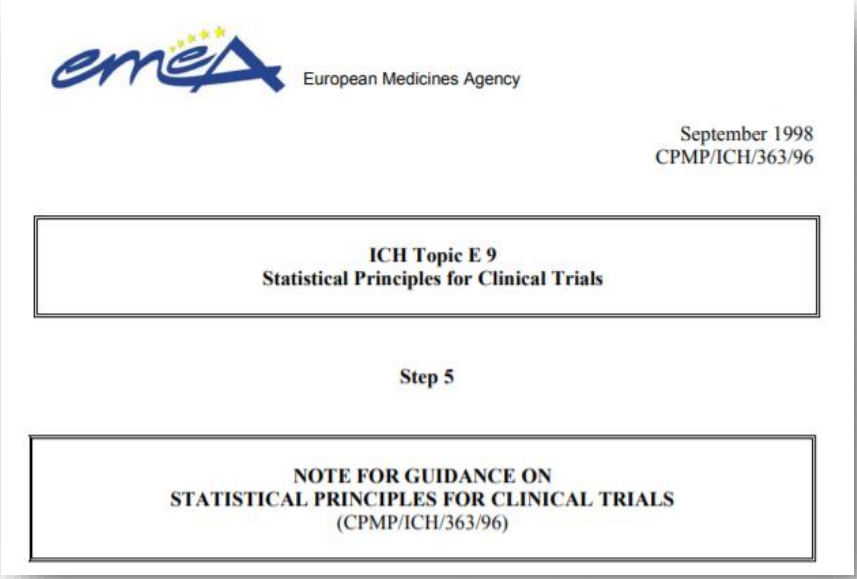
[…]

*Software verification provides **objective evidence that the design outputs of a particular phase of the software development life cycle** meet all of the specified requirements for that phase. Software verification looks for **consistency**, **completeness**, and **correctness of the software and its supporting documentation**, as it is being developed, and provides support for a subsequent conclusion that software is validated.*

The software requirements specification document should contain a written definition of the software functions. It is not possible to validate software without predetermined and documented software requirements.

Typical software requirements specify the following:

- ✓ *All software system inputs*
- ✓ *All software system outputs*
- ✓ *All functions that the software system will perform*
- ✓ *All performance requirements that the software will meet, (e.g., data throughput, reliability, and timing)*
- ✓ *The definition of all external and user interfaces, as well as any internal software-to-system interfaces*
- ✓ *How users will interact with the system*
- ✓ ***What constitutes an error and how errors should be handled***
- ✓ *Required response times*
- ✓ *The intended operating environment for the software, if this is a design constraint (e.g. hardware platform, operating system)*
- ✓ *All ranges, limits, defaults, and specific values that the software will accept*
- ✓ *All safety related requirements, specifications, features, or functions that will be implemented in software*



European Medicines Agency

September 1998
CPMP/ICH/363/96

ICH Topic E 9
Statistical Principles for Clinical Trials

Step 5

NOTE FOR GUIDANCE ON
STATISTICAL PRINCIPLES FOR CLINICAL TRIALS
(CPMP/ICH/363/96)

"The computer software used for data management and statistical analysis should be reliable[…]"

5.8 Integrity of Data and Computer Software Validity

The credibility of the numerical results of the analysis depends on the quality and validity of the methods and software (both internally and externally written) used both for data management (data entry, storage, verification, correction and retrieval) and also for processing the data statistically. Data management activities should therefore be based on thorough and effective standard operating procedures. The computer software used for data management and statistical analysis should be reliable, and documentation of appropriate software testing procedures should be available.

https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5_en.pdf

How to validate R?

- The "*Regulatory Compliance and Validation Issues - A Guidance Document for the Use of R in Regulated Clinical Trial Environments*" is available.
  https://www.r-project.org/doc/R-FDA.pdf

But this applies only to the "Base R" set.

In your everyday practice you will likely make use of numerous packages.

One of my sets →

1. ARTool
2. asbio
3. betareg
4. bindrcpp
5. binom
6. boot
7. broom
8. car
9. clubSandwich
10. compute.es
11. CRTgeeDR
12. DescTools
13. devEMF
14. dplyr
15. drgee
16. dunn.test
17. e1071
18. effectsize
19. effsize
20. emmeans
21. fitdistrplus
22. flextable
23. frailtypack
24. gee
25. geepack
26. geesmv
27. GFD
28. ggmosaic
29. ggpol
30. ggplot2
31. glmmTMB
32. gmodels
33. gplots
34. gridExtra
35. gsDesign
36. ipw
37. knitr
38. lazyeval
39. lme4
40. lmPerm
41. lsr
42. logspline
43. lunridate
44. margins
45. MCPMod
46. Mediana
47. mice
48. multcomp
49. nlme
50. Nlmixr
51. nparLD
52. officer
53. onlineFDR
54. openxlsx
55. PairedData
56. pander
57. patchwork
58. permuco
59. PK
60. PKPDmodels
61. PMCMRplus
62. PropCIs
63. qqplotr
64. quantreg
65. rlang
66. robustbase
67. robustlmm
68. RODBC
69. rstatix
70. RVAideMemoire
71. rvg
72. SASxport
73. simplexreg
74. sqldf
75. summarytools
76. survival
77. survminer
78. tidyr
79. VGAM
80. wgeesel
81. WRS2
82. xml2

What have we learned from these documents?

- Validation is a very broad term with the scope defined by the requirements
- The validation isn't about just documenting the installation or KPIs though it's important
- The validation should assess the reliability = *does it calculate correctly*?

- Before we document the installation or measure package quality (KPIs) we should first ensure that the code returns correct numbers and we can explain possible discrepancies from other (e.g. reference) software.

Who does really need it?

- Let me tell you a secret. The one who *really needs* the tool to be validated is not any agency. It is **YOU**.

- Because the one who will lose is not any agency. At the end of the day - it is

YOU

  - Statistical confirmation of the objectives is the key product of a trial
  - The final outcome is needed to approve (or reject) your drug
  - Failed trial means:
    - for you: lost money, lost reputation, lost chance (others will notice)
    - for patients: lost chance to recover, lost hopes, maybe lost lives.

Isn't this enough? Will you risk?

Is the situation really that serious? ▸ Facts

- As a matter of fact:
  - Not all key packages have exhaustive unit tests, especially those older ones.
  - Unit tests may cover only basic scenarios, limited by the author's imagination

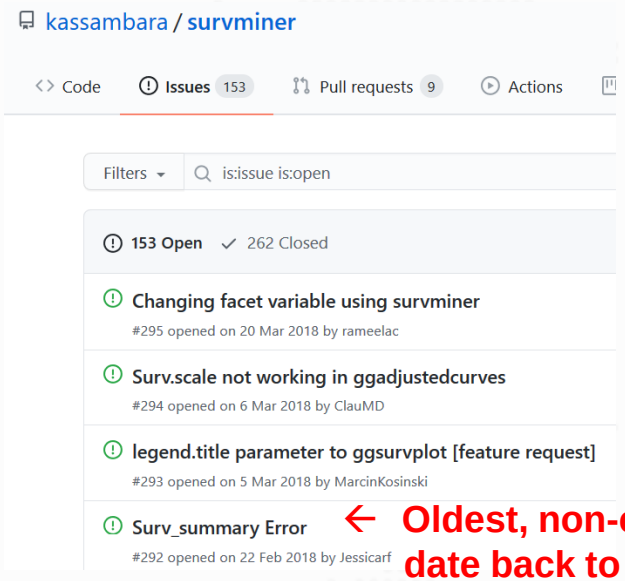
    Think about it: why are there so many "Issues" on the GitHub for the key packages, if the unit tests pass well? ← collaboration + „fresh view"
  - **There is no global authority that ensures the quality. No central QA body!**

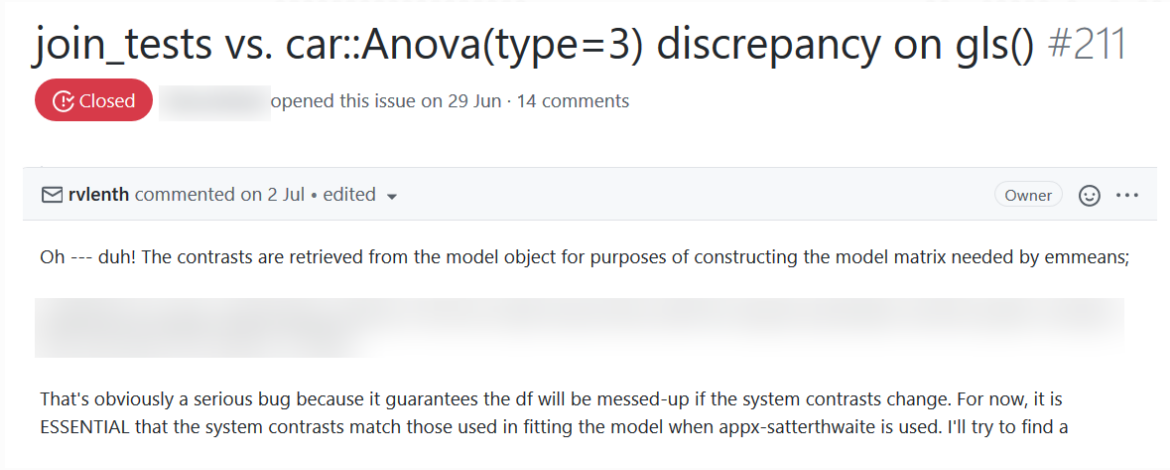
    so if anything fails, there's nobody to complain to about. *Use it at own risk!*
  - Last, but not least – and maybe the most important… **unit tests seem to be rarely subjected to comparisons against other statistical software**.

    In pharma – that's: SAS, nQuery, WinNonlin, SPSS and other tools.

- As a matter of fact:
  - Not all key packages have exhaustive unit tests, especially those older ones.
  - Unit tests may cover only basic scenarios, limited by the author's imagination

Think about it: why are there so many "Issues" on the GitHub for the key packages, if the unit tests pass well? ← collaboration + „fresh view"





← **Oldest, non-closed (maybe solved) issues date back to 2018**

Is the situation really that serious? ▸ Sad story – “No, because no”.

1 1. Compliance with SAS is not a development goal for R package maintainers. 2. Where does the documentation of the `summary.gls` function claim that robust standard errors are reported? 3. Instead the documentation says that "approximate standard errors" are provided. – Roland Jan 15 at 7:08

The problem is that

SAS *is* a major industry standard

in the Clinical Research – whether you like it or not.

Denying facts does not change the reality.

Is the situation really that serious? ▸ Facts

- So, even, if the unit tests pass well (and you trust it) – what if it does not agree with, say, SAS or WinNonlin?

  It does not mean it's wrong – maybe just differently parameterized or expressed. But in any case you should be able to explain the discrepancies, especially those noticeable ones.

  You may never realize it until asked by a reviewer, who did the calculations in another software and the results did not agree exactly.

> The statistician on my committee has assisted me in setting up my analysis design; however, he is receiving different results using SAS than I am using lmer. I understand that the goal of lmer is not to replicate results of SAS; however, I am unable to move ahead with my dissertation until I can provide him with a satisfactory explanation for these differences (I no longer have access to SAS so I need to use R as my primary statistical software).

- There is a variety of reasons for which the results generated by R may differ from corresponding outcomes obtained in other statistical packages.

Errors

Robust vs. model SEs

Origins of dates

Storage of floating point numbers

SAS: IBM, R: IEEE

Settings:

- Sum of squares
- Contrasts
- Corrections

Robust variance estimators and corrections

HC0 – HC3, CR0-CR3, CR1p,CR1S, Morell's

Degrees of freedom

Numerical issues

Rounding, numerical optimization

Random Numbers Generator

Same seed != same numbers

Algorithms:

- Quantiles
- Skewness
- Rounding

Optimization algorithm

bobyqa, Nelder-Mead, …

Estimation procedure

- There is a variety of reasons for which the results generated by R may differ from corresponding outcomes obtained in other statistical packages.

Source of discrepancy	Is problematic?	How to address it?
Errors	Yes	Fix the error or wait until fixed by the author(s) and released to the CRAN
Different algorithm: - Quantiles (9 types: SAS=3, R=S=7, 6=SPSS) - Skewness (3+ types) - Rounding. R != SPSS	No	Set appropriate option or use different method
Origins of dates	No	Just use appropriate origin
Different way of storing floating point numbers: SAS = IBM, R = IEEE	**Yes**	Nothing can be done Even simple BMI calculation (!) may give a bit different result. https://stats.stackexchange.com/questions/160711/how-to-solve-a-problem-with-different-results-in-sas-and-r-due-to-a-number-repre
Different default options, e.g. contrasts R=treatment (baseline = first) SAS = treatment (baseline = last) SPSS = deviation (sum/effect)	No (but may be very confusing)	Just set appropriate option

- There is a variety of reasons for which the results generated by R may differ from corresponding outcomes obtained in other statistical packages.

Source of discrepancy	Is problematic?	How to address it?
Different “pshilosphy”, e.g default type of sum of squares (available out-of-the-box) – “holy wars” R – type I (sequential) SAS – type III (marginal)	No	Use appropriate package (aov, car::Anova, anova(type=xx), emmeans::joint_tests)
Differences in random number generators Same seed = different numbers	Yes	Nothing can be done (maybe there are packages with the same RNG as in SAS)
Different optimizing method	No / Yes	No – if the same method can be set in both packages and the results agree. Otherwise nothing can be done (without implementing it)
Differences in estimation method	No / Yes	No – if the same method can be set in both packages. Otherwise nothing can be done (without implementing it)
More complex settings, like the type of robust variance estimator: HC0 – HC3, CR0-CR3, CR1p,CR1S, Morell’s correction for small sample	No / Yes	No, if both methods allow to set this option. In R this is spread multiple packages.

- There is a variety of reasons for which the results generated by R may differ from corresponding outcomes obtained in other statistical packages.

Source of discrepancy	Is problematic?	How to address it?
"Big stories" – mixed models: the way the degrees of freedom are calculated, estimation method, optimization method, standard errors, dealing with both random (G) effect and residual (R) covariance structures. SAS – PROC MIXED, GLIMMIX R: glmmPQL, glmmTMB, lmer, lme, gls (=MMRM with REPEAT), MCMCglmm **Neither of** (the frequentist) **R packages can do what SAS does**. The glmmPQL and nlme are useful in longitudinal analysis (including the MMRM). Satterthwaite method for DFs is available for both lme4 and nlme (simulated). Kenward-Roger - only for lme4. lme4 doesn't handle R+G covariance at the same time. GLM is handled by glmmPQL (biased), lme4 and glmmTMB (only Wald's tests and no KR/Satt.).	Yes, very	In more complex scenarios there may be **no way to obtain the same results in R and SAS**, so there is no way to validate the calculations exactly.

- But differences may occur even within R itself.

  R is well known for having multiple implementation of the same method, which confuses the users if the results differ. A few examples:

  - Mixed models – as mentioned previously
  - anova vs. car::Anova
  - confidence intervals: normal vs. bootstrapped implementation (BCa, percentile, studentized, parametric)
  - sample size for the log-rank test: Schoenfeld vs. Freedman
  - different optimization and estimation algorithm
  - t vs. z
  - LS-means vs. raw means

How to validate R?

- I guess you are now **convinced**, that using R “out-of-the-box”, without thorough numerical validation may be a dangerous idea.

  Even, if the used routine is correct, the discrepancies from SAS (or other software) may be noticeable and **a reviewer may ask you to explain it**.

Statistical Reviewer

```
Call:
glm(formula = y ~ x1 + x2, family = Gamma(link = "log"), data = test)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-0.26213  -0.08456  -0.01033   0.08364   0.20878

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.9210757  0.0324674 213.170   <2e-16 ***
x1          -0.0003371  0.0005985  -0.563    0.575
x2           0.0234097  0.0627251   0.373    0.710
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’

(Dispersion parameter for Gamma family taken

    Null deviance: 1.3498  on 99  degrees of
Residual deviance: 1.3436  on 97  degrees of
AIC: 1240.6

Number of Fisher S
```

https://stackoverflow.com/questions/44577998/standard-errors-discrepancies-between-sas-and-r-for-glm-gamma-distribution

```
summary(fit, dispersion=1/gamma.shape(fit)$alpha))
```

the summary function will use the MLE of alpha when computing the SEs, and they will match SAS/genmod exactly.

Analysis Of Maximum Likelihood Parameter Estimates

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	6.9211	0.0323	6.8577	6.9844	45836.1	<.0001
x1	1	-0.0003	0.0006	-0.0015	0.0008	0.32	0.5718
x2	1	0.0234	0.0610	-0.0962	0.1430	0.15	0.7013
Scale	1	74.5945	10.5258	56.5714	98.3596		

Is the situation really that serious? ▸ Discrepancies from other software – part 3a

Differences between PROC Mixed and lme / lmer in R - degrees of freedom

Asked 8 years, 4 months ago Active 8 years, 4 months ago Viewed 11k times

12

Note : this question is a repost, as my previous question had to be dele

While comparing PROC MIXED from SAS with the function `lme` from th
stumbled upon some rather confusing differences. More specifically, the
different tests differ between `PROC MIXED` and `lme`, and I wondered why

Is the algorithm of calculating SEs of beta coefficients calculated by the nlme gls finally fixed?

Asked 8 months ago Active 8 months ago Viewed 57 times

4

I can read here: https://www4.stat.ncsu.edu/~davidian/st732/examples/dental_pa.R and here: https://math.unm.edu/~luyan/stat57918/week14.pdf

that:

> WARNING: There is a MISTAKE in gls(), and it DOES NOT calculate the model-based sampling covariance matrix of betahat correctly! Thus, the model-based standard errors that gls() reports are ever so slightly "off" from the correct values. MOREOVER, gls() does not calculate the robust sandwich covarianc
> robust (or empirical) standard errors. The fu
> model-based standard errors along with rob

Upcoming Events

2020 Community Moder
ends Oct 20

Featured on Meta

Goodbye, Prettify. Hello
Swapping out our Syntax

3

JMP centers polynomials by default. You can override that default by uncli
the red triangle by "Model Specification" in the "Fit Model" dialog box. Doin
same results for the Type III sums-of-squares as R, when you use the car
contrasts. In other words, if you wish to reproduce JMP's default results in
numerical predictor in R first. If you wish to reproduce R's default in JMP, y
"Center Polynomials" option in JMP first.

- SAS and R use different conventions to calculate AIC and BIC. We have used ML here, in which case AIC is the same but BIC differs as noted above.
- If we'd used REML, both AIC and BIC values are calculated differently by SAS and R using different conventions regarding the number of observations and number of parameters, so are not comparable, but can be compared within a single

Cox proportional hazard model in R vs Stata

Asked 4 years, 10 months ago Active 4 years, 10 months ago Viewed 1k times

I'm trying to replicate in R a cox proportional hazard model estimation fr

Model (a): unstructured Σ_i

- Common unstructured correlation Σ_i with variances changing over time for both genders
- Note that gls() defines BIC dfferently from SAS (it uses the total number of observations N while SAS MIXED uses the total number of individuals m
- The weights statement makes the variances on the diagonal differ over time - the default with no weight statement is that they are the same for all times
- Note that standard error estimates of $\hat{\beta}$ are not correct, need to use robust.cov function to derive the correct ses.

Note

The p-value differs slightly from that of SAS because a second order term is included in the asymptotic approximation in R.

References

T. W. Anderson (1958). *An Introduction to Multivariate Statistical Analysis.* Wiley.

> *Second, be sure to understand that reproducing a SAS analysis with lme in*
> *no way violates any legal agreements that SAS may have, if for no other*
> *reason than you never signed an agreement with SAS! That bit in the EULA*
> *about decompiling and reverse engineering means that people are prohibited*
> *from creating a new version of PROC MIXED that does the same thing. The*
> *nlme package uses different methods than SAS. E.g. different optimizers,*
> *even uses a log-parameterization deep in the code so that negative variance*
> *components cannot happen.*
>
> *Third, by now you've probably figured out that PROC MIXED and lme have very*
> *different ideas about degrees of freedom. Also, the loglikelihoods are on*
> *different scales. For that reason, when I try to reproduce an analysis, I*
> *find the best way to compare is to look at the variance components.*

Is the situation really that serious? ▸ Discrepancies from other software – part 4a

Why do anova(type='marginal') and anova(type='III') yield different results on lmer() models?

Asked 1 year, 6 months ago Active 10 months ago Viewed 574 times

3

When analysing mixed-effects data using lmer() I find that using anova(type='marginal') and anova(type='III') give different results. Why the discrepancy?

The results from anova(type='marginal') are identical to those I get from using car:Anova(type='III') on the same model and to using both anova(type='marginal') and car:Anova(type='III') c
data fitted using lme().

1

Upcoming Events

2020 Community Moder
ends Oct 20

Why do lme and aov return different results for repeated measures ANOVA in R?

Asked 9 years, 2 months ago Active 4 years, 5 months ago Viewed 19k times

24

I am trying to move from using the `ez` package to `lme` for repeated measures ANOVA (as I hope I will be able to use custom contrasts on with `lme`).

Following the advice from this blog post I was able to set up the same model using both `aov` (as does `ez`, when requested) and `lme`. However, whereas in the example given in that post the *F*-values do perfectly agree between `aov` and `lme` (I checked it, and they do), this is not the case for my data. Although the *F*-values are similar, they are not the same.

Upcoming Events

2020 Community Modera
ends Oct 20

Featured on Meta

[R] difference between coxph and cph

Wed Apr 21 17:07:57 CEST 2004

- Previous message: [R] difference between coxph and cph
- Next message: [R] difference between coxph and cph
- **Messages sorted by:** [date] [thread] [subject] [author]

```
On Wed, Apr 21, 2004 at 07:09:03AM -0400, Frank E Harrell Jr wrote:
[...]

> No, cph is essentially a wrapper for coxph and uses the same computations.
>  The problem is that Deb did not read the documentation to summary.Design
> nor the Overview of the Design package.

And obviously that I didn't :-(

[...]
> The Design package by default computes inter-quartile-range effects.
[...]

Interesting idea... is this a common procedure in biostatistics?
```

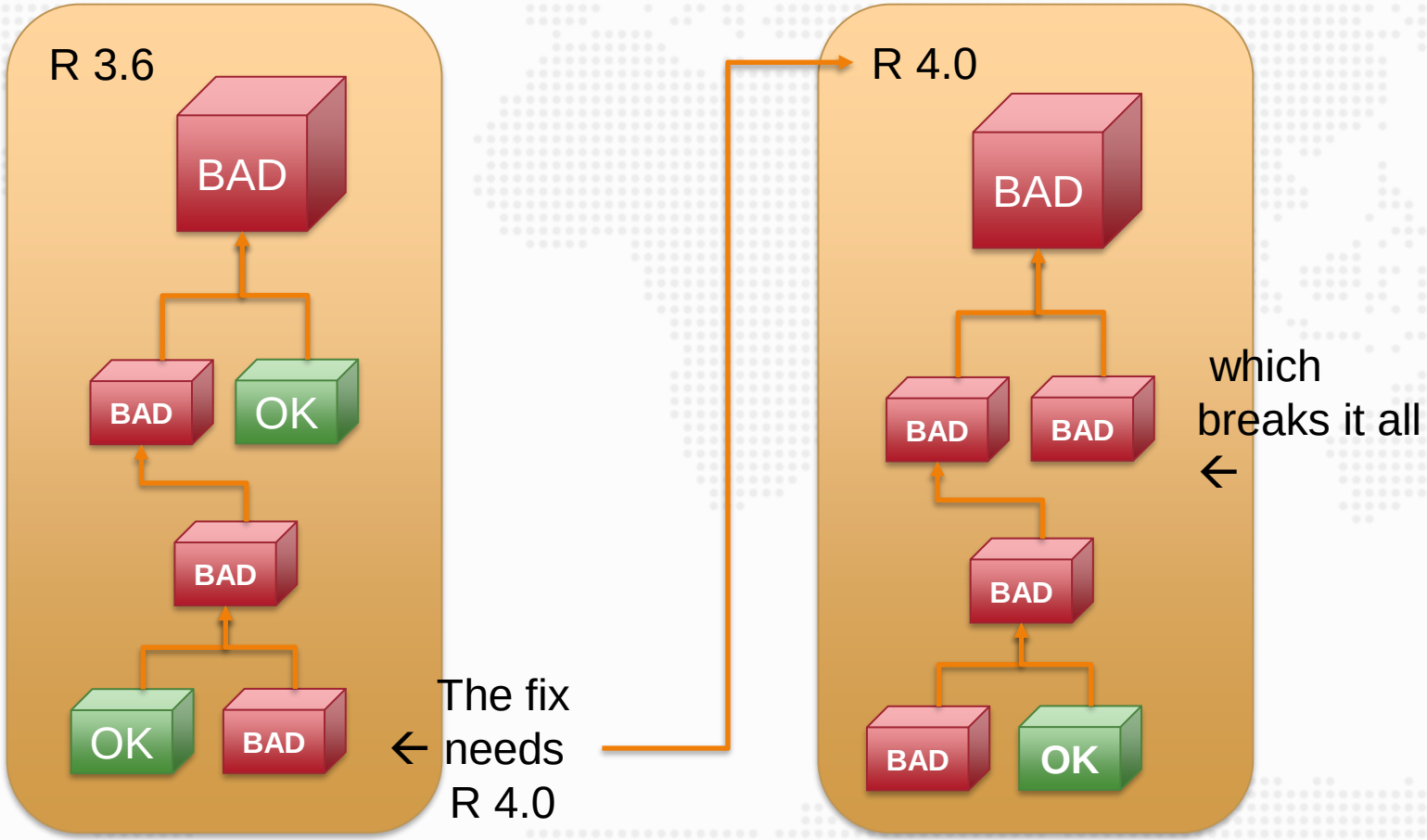
How to validate R?

- How can R be validated numerically?
  - By comparing the outcomes to the output of a reference software (e.g. SAS, SPSS, Stata). It requires the access to the software or asking someone who can do the calculations for us.
  - As above - by using examples, with attached data sets and results, published in software manuals (SAS, NCSS)
  - By comparing the outcome with another function in R that is already validated
  - By inspecting the R source code and comparing it with textbooks formulas
    This works only for the simplest cases, like the t.test().
    More complex routines often employ advanced **numerical optimization**.
    For example: $\mathbf{(X^TX)^{-1}X^Ty}$ vs **QR** factorization, eigenvalue decomposition vs. SVD

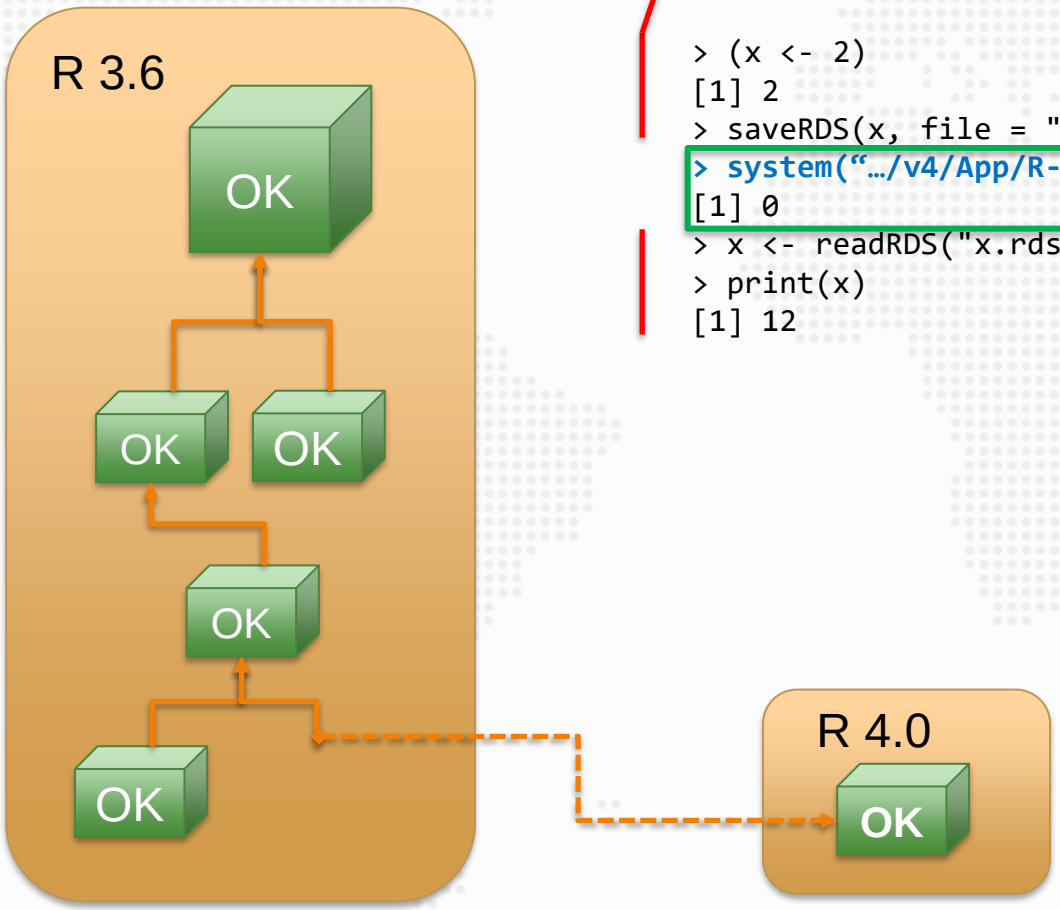
- When will you decide that the outcomes agree and the function is validated?
  - Exact agreement SAS: 10.21 vs. R: 10.21 (but mind the floating point issue!)
  - Agreement to n-th decimal place SAS: 10.211 vs. R: 10.215
  - Agreement "about the same" SAS: 10.211 vs. R: 10.376
  - Agreement to the order of magnitude SAS: 10.2 vs. R: 14.6
  - Problem with methods using sampling - no exact comparison (for the same seed)
  - Problem at boundaries: p-values (SAS: 0.048, R: 0.052), CIs (includes 0/1 or not), bayesian factors – and any other threshold used for making binary decisions
  - Partial agreement: function X returns two outcomes A and B. For A you get exact agreement, for B – only partial. What is the status of the function?

- Problems
  - R is a very dynamic ecosystem. Packages are updated frequently, most recent bug fixes are published on the GitHub rather than CRAN
  - Packages are mutually dependent. Each unverified dependency breaks the chain
  - New releases (with bug fixes) may require upgrade other packages or even the R core itself (!) – which may result in upgrade of all installed packages, which eventually may break the existing code.
  - No central authority that validates packages is (yet) available – each CRO has to do it on one's own. The same work has to be repeated over and over.
  - Packages evolve dynamically. What is legal today, tomorrow may be obsolete or removed without a warning. In this case – all unit tests will fail, if depending on it.

"Quick fix" may require an upgrade of the R core, which may invalidate the previously validated and working code.



Portable R (in Windows)



R 3.6 – most of the analysis, R 4.0 – fixed routine

```
> (x <- 2)
[1] 2
> saveRDS(x, file = "x.rds")
> system("…/v4/App/R-Portable/bin/x64/Rscript.exe myscript.r")
[1] 0
> x <- readRDS("x.rds")
> print(x)
[1] 12
```

myscript.r

```
x <- readRDS("x.rds")
x <- x + 10
saveRDS(x, "x.rds")
```

Rule: do the most of the analysis in the most tested version of R (here: 3.6), call upgraded R (here 4.0) only for certain tasks – easier to manage

Portable R (in Windows)

How to validate R? ▸ Obstacles 2

Package 'StatCharrms' was removed from the CRAN repository.

Formerly available versions can be obtained from the archive.

Archived on 2020-10-02 as check issues were not corrected in time.

The most recent check results can be obtained from the check results archive.

Please use the canonical form https://CRAN.R-project.org/package=StatCharrms to link to this page.

R packages are "live", they change with time. Form version 2.44 of the package "survival", summary.coxph does not report R2 anymore. You can read the reason for this in the "Changes in version 2.44":

> The Nagelkirke R^2 has been removed from summary.coxph. The shortcomings measure are well known, concordance is a better measure.

If your preferred package(s) switch(es) from data.table to tibbles →

```
> (x <- data.frame(A=1:3, B=1:3))
  A B
1 1 1
2 2 2
3 3 3
> x[, 1]
[1] 1 2 3
> dim(x[, 1])
NULL
> length(x[, 1])
[1] 3
>
> x <- as_tibble(x)
> x[, 1]
# A tibble: 3 x 1
      A
  <int>
1     1
2     2
3     3
> dim(x[, 1])
[1] 3 1
> length(x[, 1])
[1] 1
>
```

```
Note: Using an external vector in selections is ambiguous.
i Use `all_of(stat_choice)` instead of `stat_choice` to silence this message.
i See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.
This message is displayed once per session.
# A tibble: 12 x 3
```

"a few" warnings from the dplyr – it's valid code written a few years ago.

```
Warning messages:
1: `funs()` is deprecated as of dplyr 0.8.0.
Please use a list of either functions or lambdas:

  # Simple named list:
  list(mean = mean, median = median)

  # Auto named with `tibble::lst()`:
  tibble::lst(mean, median)

  # Using lambdas
  list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
This warning is displayed once every 8 hours.
Call `lifecycle::last_warnings()` to see where this warning was generated.
2: The `...` argument of `summarise_at()` can't contain quosures. as of dplyr 0.8.3.
Please use a one-sided formula, a function, or a function name.
This warning is displayed once every 8 hours.
Call `lifecycle::last_warnings()` to see where this warning was generated.
>
```

kassambara commented on 30 Jul 2019

Many users have experienced an issue when trying to install the latest survminer version (v 0.4.4) from CRAN.

This is due to the recent update of the package `cmprsk`, which suddenly requires the current R version >= 3.6.0, forcing survminer users to update their R version.

◉ Considerations

- The numerical validation **consumes time** (=money), and needs special efforts.

- Fortunately, it is an incremental process. Only the **used functions have to be validated** (not all available ones!). That is – the smallest part of the validation is a **package::function** part, not the entire package itself (possibly exposing numerous functions you may never require)

- Once test cases are prepared, they can be stored into a repository and run as needed. The **library grows over time**, utilizing data from new trials.

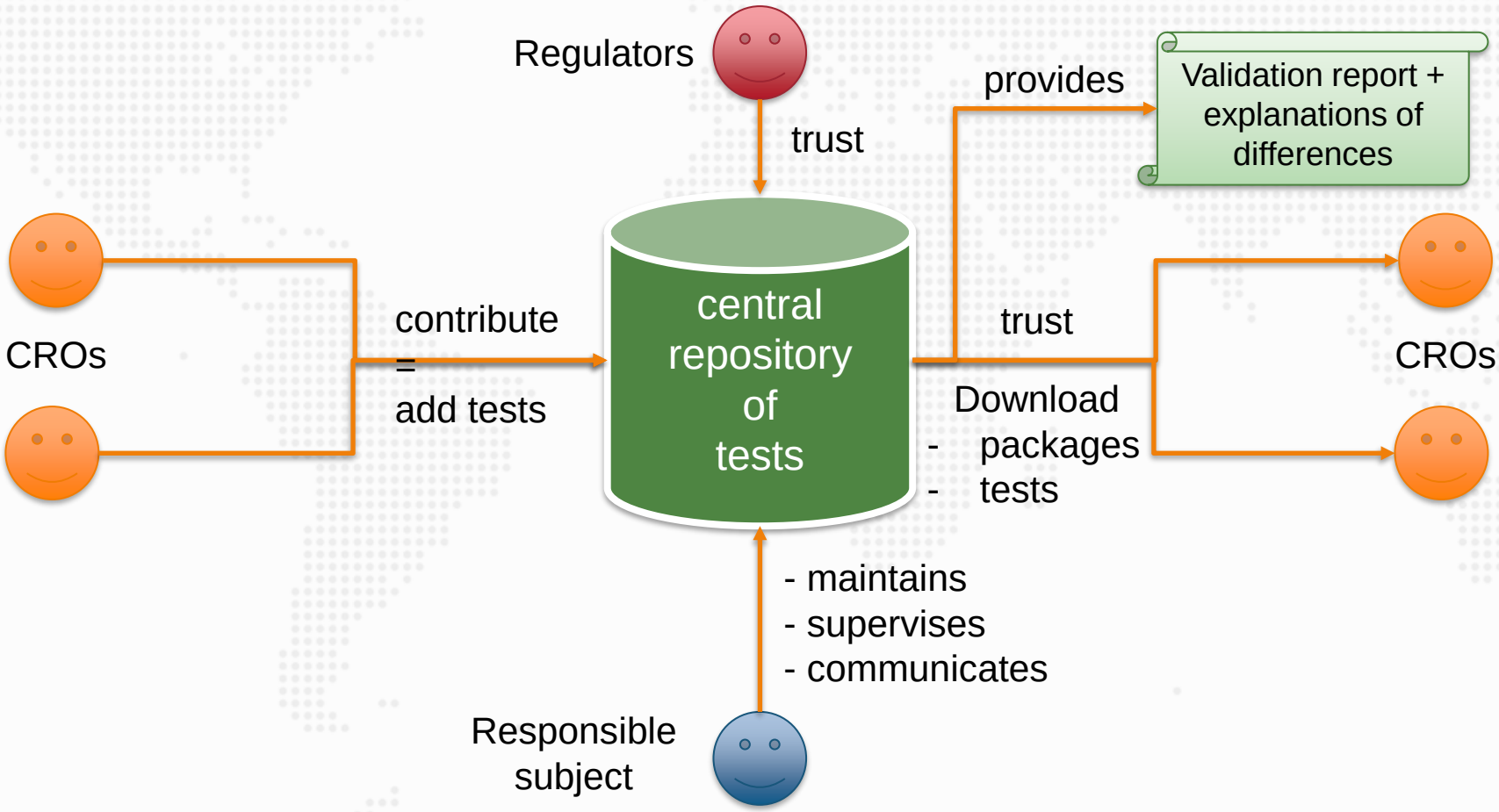
- **It is doable** by a single person but only assuming a good availability of resources and reasonable time to spend (several months).

What would help?

- A global project on the validation of R, that:
  - is trusted by both the regulators and the industry
  - is collaborative – anyone can send verified ***testthat*** cases (with data), if well documented, with attached print out of the reference software
  - Is easy to use by the departments of biostatistics (or IT)
  - Reacts quickly to changes! R changes much faster than SAS, inertia here means outdated packages, unfixed errors, missing functionalities. Possibly it could offer tools to validate it on one's own (ad hoc validation in urgent cases)
  - Provides a way to quantify the results of validation (as mentioned before)
  - Provides explanation to all discrepancies from at least SAS (ideally also other counting software)

It could be fundraised, grant-based, donated, paid-per-subscription. Subjects who contribute the most could get the access for free.

What would help?



What would help?

- Such project would enable **much wider use of R in submissions** of demanding RCTs, where the risk of potential serious problems, maybe even leading to general failure of the trial, **may prevent the managers from considering R a safe, reliable option** or ever the replacement for SAS.

  That's one of the reasons for "*… **they hesitate***".

- The presented idea seems to differ from the idea of the ***R Validation Hub***, at least currently (https://www.pharmar.org). Both may nicely complement each other, though.

Other approaches ▸ a warning

- With the “KPI” approach, one should be very cautious about the measures like:
  - **package popularity**; the ***nlme*** package is rather “unpopular” compared to the lme4, while being **the core of the MMRM** (one of the key models in CR)
  - **availability of vignettes, websites and NEWS, GitHub**; the ***nlme*** package has no: vignette, website, NEWS, GitHub repo (other than r/o) – only the changelog.
  - **frequency of updates** – stable and “conservative” packages may be updated infrequently. Frequent updates don’t necessarily correlate with key importance.

→ *nlme would receive poor score and not pass criteria for being „recommended”*

```
2020-08-20 Peter Dalgaard

        * NAMESPACE, R/corStruct.R, man/corFactor.corStruct
          man/corMatrix.corStruct.Rd, tests/corFactor.R: Ap
          from Sebastian Meyer to fix misnamed
          corFactor.compSymm -> corFactor.corCompSymm PR#16
```

```
2019-01-23  Martin Maechler  <maechler@stat.math.ethz.ch>

        * R/lme.R (lme.formula, lmeControl): new option
        'allow.n.lt.q=FALSE' by default now triggers error
        "fewer observations than random effects in all level <Q> groups".

        * src/nlmefit.c (finite_diff_Hess): prevent integer overflow (and
        later seg.fault) for large 'nTot' (already for npar >= 305).
```

Summary

- Numerical validation of R is **important for YOUR safety**
- The nature of R differs from the conservative nature of SAS. Things are scattered across packages and versions. **The R ecosystem is dynamic**.
- Discrepancies with other software occur quite often
- Some of them may be easy to address and explain, but some indicate errors. Do not ignore them. Don’t assume your tool is right (and the other is wrong)
- Numerical validation is totally doable but consumes time
- There is a need for central, trusted, collaborative repository of unit tests

This is just the beginning…

THANK YOU