

Using R to Validate Results of Other Programming Environments

Nikhil Abhyankar, Vidyagouri Prayag
Cytel

IS06 - PhUSE 2011

Disclaimer

This presentation is based on real professional experience of the presenters, their opinions and learning

However this is not to be construed as an official statement endorsed by Cytel

Agenda

- What FDA wants and likes
- Why Validation?
- Use of R at Cytel
 - Software Development
 - Clinical Trial Data Reporting
- Strengths of R
- Limitations and challenges
- Conclusions

What FDA wants and likes

- Safety of subjects
- Results to be accurate
- Results to be reliable and reproducible
- Amenable to a review

Why Validation?

- To ensure high quality of the software/deliverable
- To ensure reliability of the product
- Compliance with specified requirements
- Assurance of accuracy – a double check
- Test for reproducibility of results

Independent Validation

- Parallel and independent workflow for development and validation
- Distinct teams working on development and validation
- Different software / programming tools used by the teams
- More alert more reliable, no copy – paste job!

What value do you receive?

Confidence
In delivering what is required!

Need Validation for

- Software Development

Complex and innovative algorithms programmed in C++

- Clinical Trial Data Analysis and Reporting

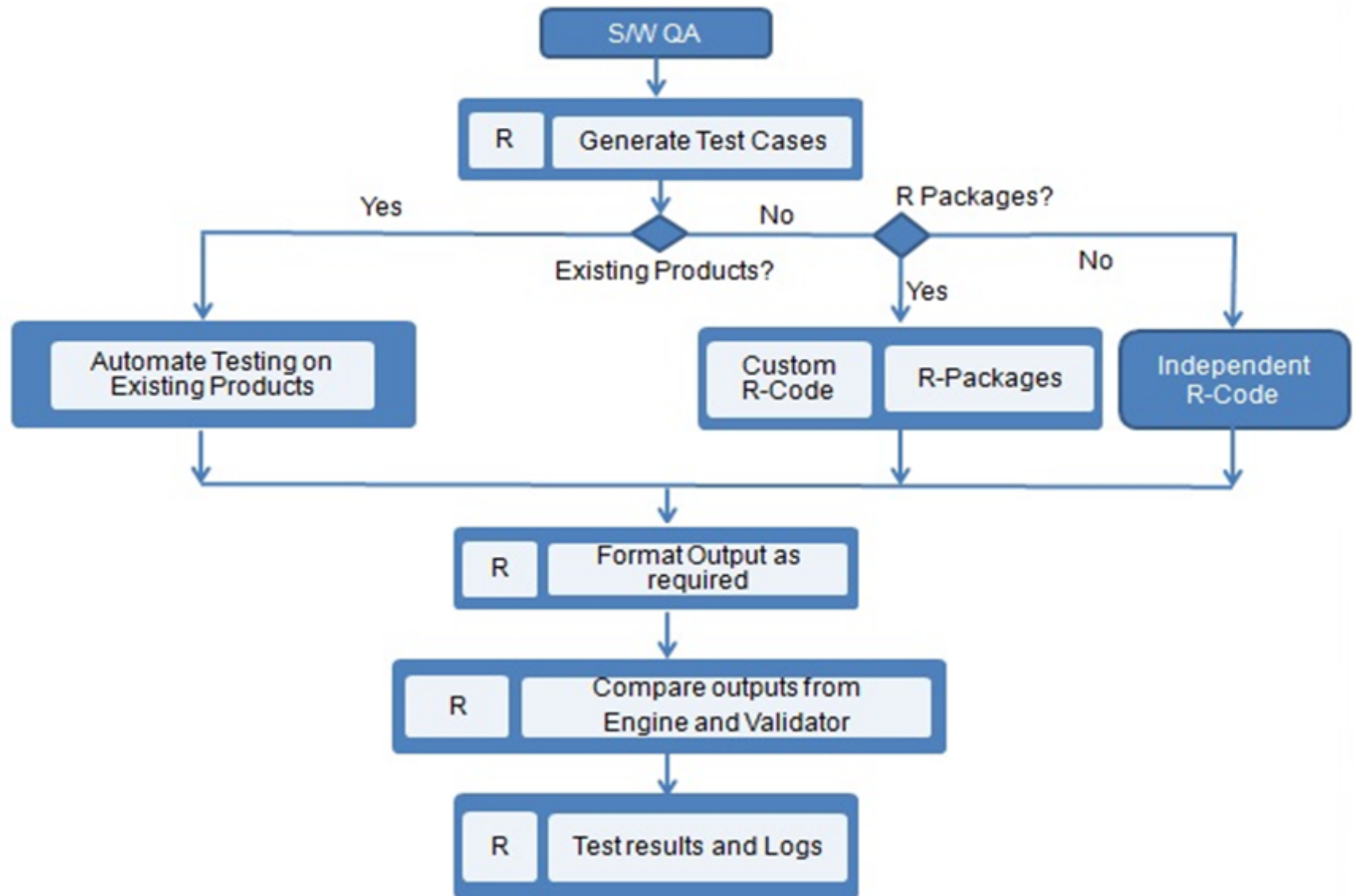
Clinical data reports generated using SAS®

Validating 'First-to-market' software?

Developing the same feature in another programming language is ***the*** validation

R is a great choice!

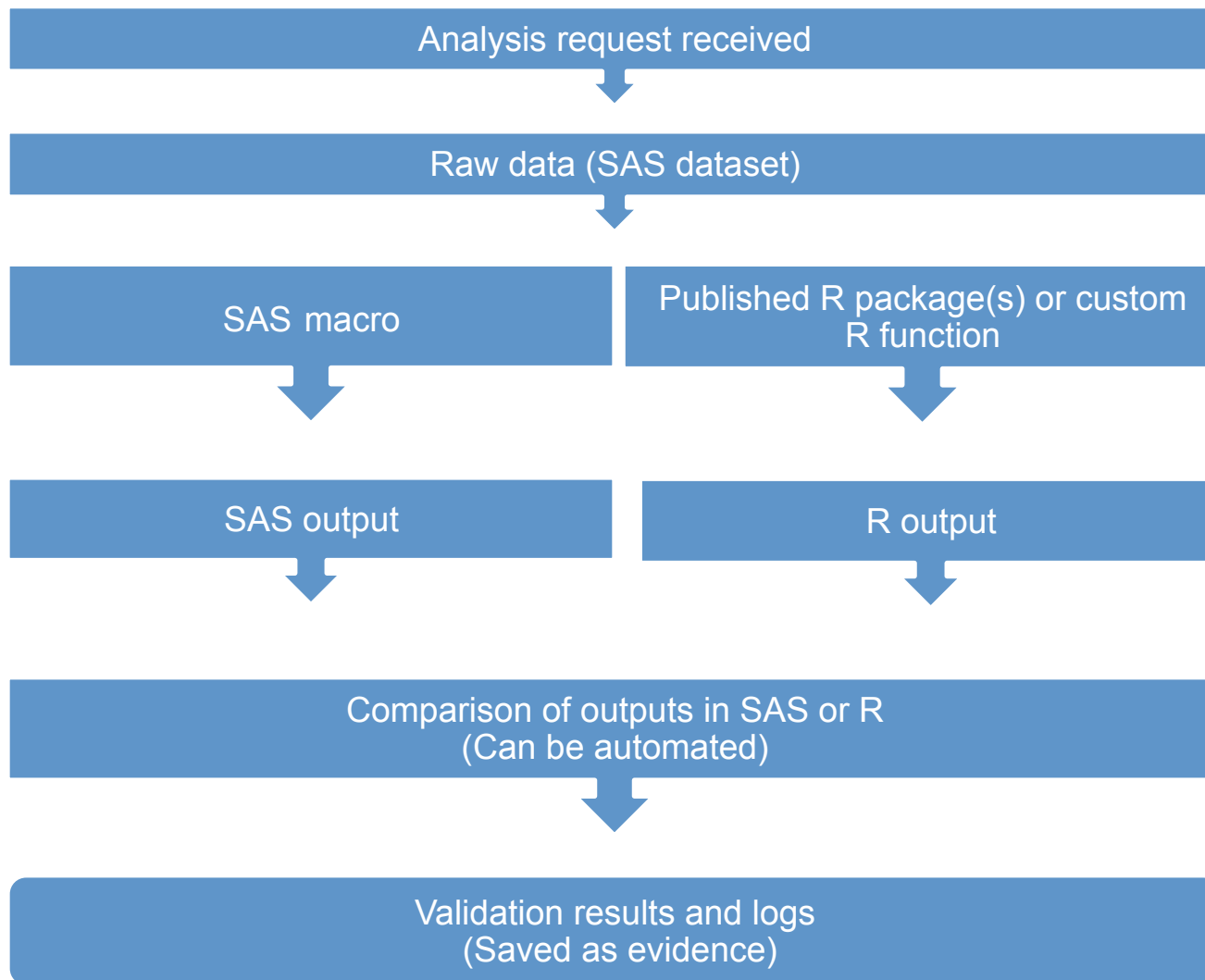
Software Quality Assurance @ Cytel



Software Quality Assurance @ Cytel

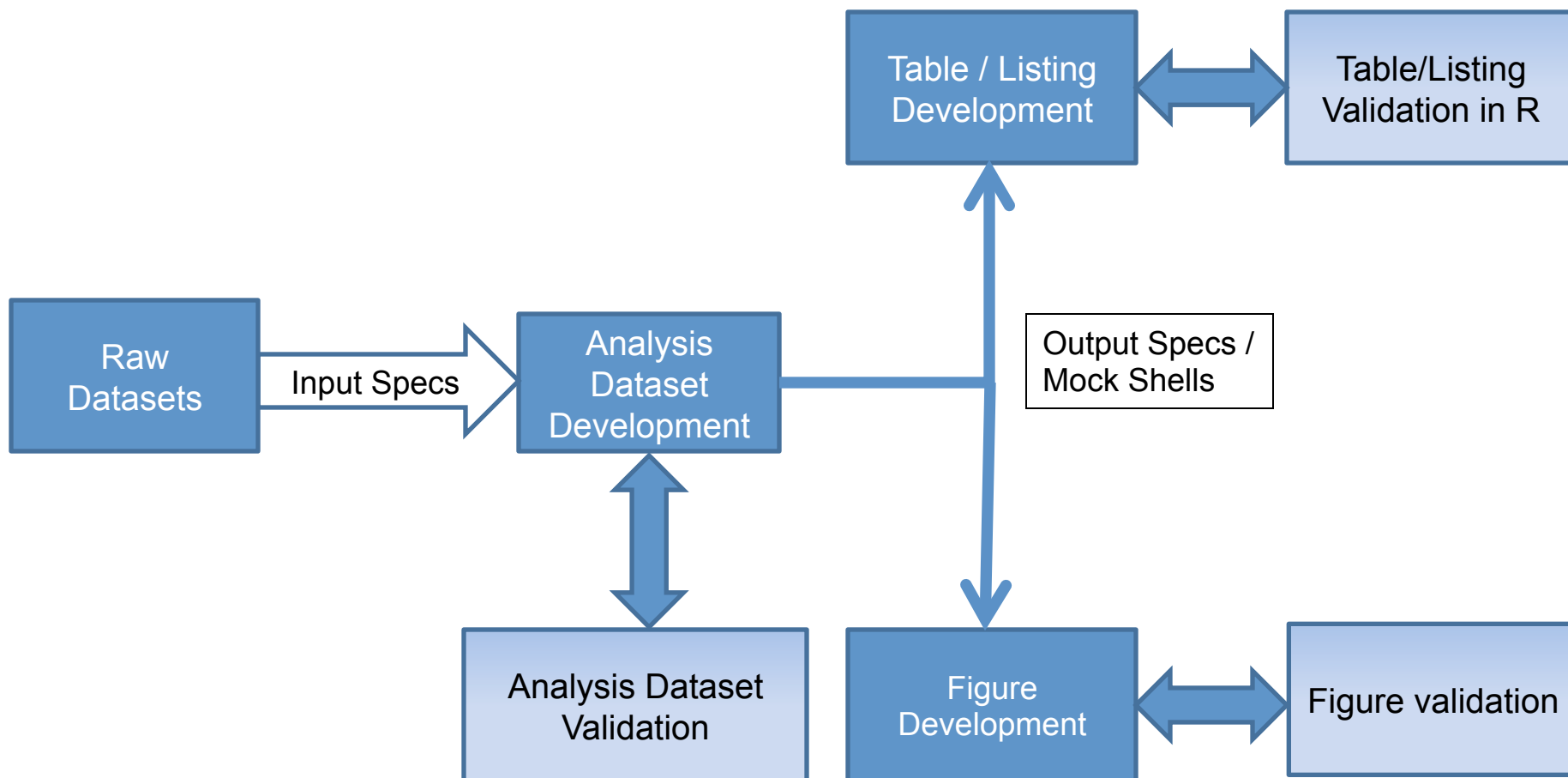
- Generation of test cases and automation
- Intermediate output / internal consistency checks
- Graphical validation for simulation output
- Improving the development of the algorithm
- Calling R from products

Clinical Data Analysis and Reporting @ Cytel



Clinical Data Analysis and Reporting @

Cytel



Clinical Data Analysis and Reporting @

Cytel

R function for safety reports

(ECG, Vital signs and Lab tests)

- Input SAS data and exploratory analysis
- Formatted time-points
- Summary statistics

Clinical Study Reports (CSR) and PK, PD Reports

- Statistical Sections of CSR
- Subject listings
- Model based summaries

Case for the 'other'

- The data:

	seed	subject	dose	response
1	689153326	1	50	-0.326592916
2	1945691870	2	10	1.5424368654
3	474838741	2	20	0.2590291308
4	854874385	2	30	-0.555833885
5	403052287	2	40	0.7787203766
6	936462086	3	10	-0.655197048
7	566297764	3	20	-0.022104952
8	1191547602	3	30	-0.772646961
9	1849723452	3	40	0.7242335814
10	1847748147	4	10	1.2633145661
11	1653466969	4	20	0.1154359593
12	705985501	4	30	-0.39933183
13	2004457304	4	40	1.0417314406

Case for the 'other'

- Regression – proc reg (SAS) and lm (R)
 - Estimated dose level with sample of size 1 for a by group

SAS Code:

```
proc reg data = phuse outest = out noprint;  
    model response = dose;  
    by subject;  
run;
```

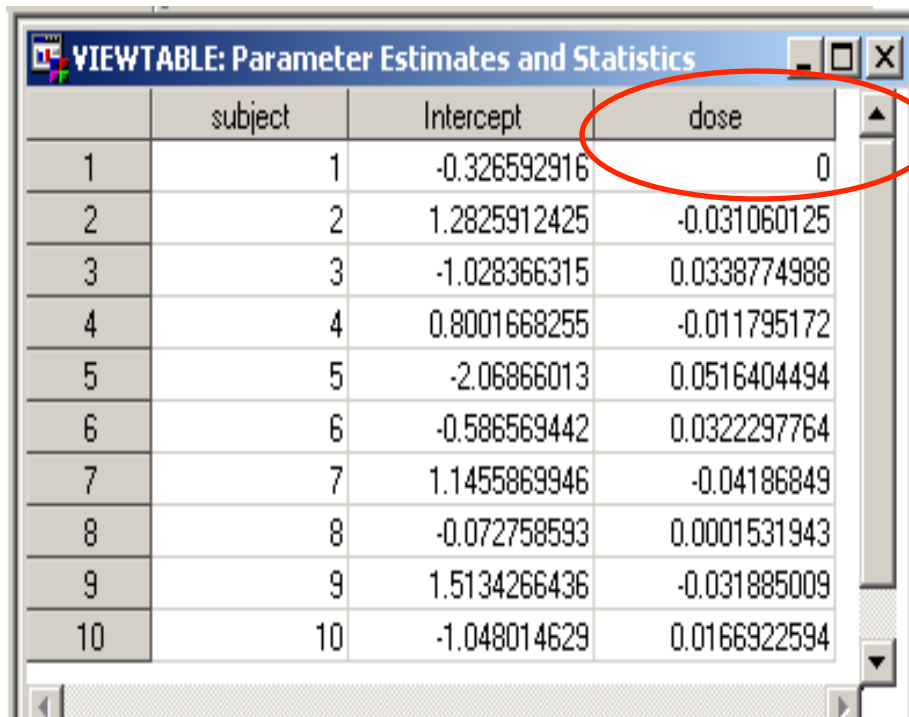
R Code:

```
> library(nlme) # Invoking the package 'nlme'  
> coeff_icept = as.vector (by (data = phuse, subject, function(x)  
    lm (response ~ dose, data = x)$coefficients [1]))  
    # Saving the intercepts  
> coeff_dose = as.vector (by (data = phuse, subject, function(x)  
    lm (response ~ dose, data = x)$coefficients [2]))  
    # Saving the estimates for the dose
```

Case for the 'other'

The outputs showing the estimate of dose to be NA or zero for subject =1

SAS



	subject	Intercept	dose
1	1	-0.326592916	0
2	2	1.2825912425	-0.031060125
3	3	-1.028366315	0.0338774988
4	4	0.8001668255	-0.011795172
5	5	-2.06866013	0.0516404494
6	6	-0.586569442	0.0322297764
7	7	1.1455869946	-0.04186849
8	8	-0.072758593	0.0001531943
9	9	1.5134266436	-0.031885009
10	10	-1.048014629	0.0166922594

R

	coeff_icept	coeff_dose
1	-0.32659292	NA
2	1.28259124	-0.031060125
3	-1.02836631	0.033877499
4	0.80016683	-0.011795172
5	-2.06866013	0.051640449
6	-0.58656944	0.032229776
7	1.14558699	-0.04186849
8	-0.07275859	0.000153194
9	1.51342664	-0.031885009
10	-1.04801463	0.016692259

Another case for the 'other'

With thanks from:

<http://stackoverflow.com/questions/33916>

[r](#)

[sas-and-](#)

	x1	x2	x3	y
1	2.66	20	0	0
2	2.89	22	0	0
3	3.28	24	0	0
4	2.92	12	0	0
5	4	21	0	1
6	2.86	17	0	0
7	2.76	17	0	0
8	2.87	21	0	0
9	3.03	25	0	0
10	3.92	29	0	1

- SAS Code

```
proc qlim data = probit;
```

```
model y = x1 x2 x3/ discrete  
      (d=probit);
```

```
output out = outdata marginal;
```

```
run;
```

```
quit;
```

- R Commands

```
> model <- glm(y ~ x1 + x2 + x3,  
               family = binomial (link="probit"),  
               data = probit)
```

```
> summary (model)
```

Another case for the 'other' – the differing outputs

(SAS) Parameter Estimates

Parameter	Estimate	Standard Error	t Value	Approx Pr > t
Intercept	-7.452320	2.542473	-2.93	0.0034
x1	1.625810	0.693882	2.34	0.0191
x2	0.051729	0.083890	0.62	0.5375
x3	1.426332	0.595038	2.40	0.0165

(R) Coefficients

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-7.45231	2.57152	-2.898	0.00376
x1	1.62581	0.68973	2.357	0.01841
x2	0.05173	0.08119	0.637	0.52406
x3	1.42633	0.58695	2.430	0.01510

Keeps the validator alert and on the toes!

Strengths of R

- Free Free Free!



- Flexible data acrobatics
- Bleeding edge developments - numerous packages implementing the latest in theory
- Updating your favorite packages in R is an easy task
- R allows the use of variables from different 'data frames' or other data structures for analyses and graphics
- Easy to add user defined functions in R

Limitations and challenges

- Help is technical, less accessible to a non techie
- Need to be aware of latest available packages and functions
- Open source R is not the best in handling huge datasets and complex looping
- Market perception to disregard free software as not being up to the mark

Conclusions

- Use of another software recommended for validation
- R is a great choice for validation of both, statistical software and analysis
- R can be used for validation of routine and advanced analysis as well as innovative, first to market products

References

- <http://www.r-project.org/>
- <http://cran.r-project.org/web/packages/SASmixed/vignettes/Usinglmer.pdf>
- <http://www.springer.com/statistics/computational+statistics/book/978-1-4614-0684-6>
- <http://www.r-project.org/doc/R-FDA.pdf>
- <http://cran.r-project.org/web/packages/digitize/index.html>

Looking forward...

From



to



Questions, doubts?



nikhil.abhyankar@cytel.com
vidya.prayag@cytel.com