

An Efficient Way for Statistical Review Using the RShiny Application

Adarsh Nagare, Cytel, Pune, India

Sameer Bamnote, Cytel, Pune, India

ABSTRACT

Statisticians often need to review critical reports before study submissions. Many reports include outputs based on statistical testing or model-based analysis. To ensure the accuracy of these reports, most of the time the Statistician needs to write independent code. In case of tight timelines, it may not be feasible and quality of review can be impacted.

To address this we describe use of an application that can save time for the Statistician. This application was created using open-source software R and Shiny, providing an efficient, time saving solution for review of critical statistical outputs without the need for extensive code writing. The aim is to review results for t-tests, AN(C)OVA, their non-parametric alternatives and model-based analysis which are performed in SAS[®]. User has flexibility to provide input data in excel, csv or SAS format. User can select the statistical test (or specify the model) as per requirement and statistical outputs will be just a click away. This paper demonstrates the application with a variety of illustrations.

INTRODUCTION

The authors work with clients from clinical research and FMCG field as Statisticians. As a part of the role we have to deal with programmers routinely to do the desired work (i.e. Tables, Listings and Figures) required for various reports. Most of the time, as a Statistician we need to validate the results for modelling or statistical tests or cross check them by writing independent code to ensure accuracy of the results, prior to the submissions. The usual practice is to use SAS to validate the results, which is time consuming. This might not be feasible every time because of tight timelines.

To handle this situation authors developed an application in R together with Shiny. This application focuses on various statistical tests like t-tests, AN(C)OVA, their non-parametric alternatives like Wilcoxon Signed Rank Test, Wilcoxon Rank Sum Test, Kruskal-Wallis Test etc. and model-based analysis like mixed effect model with and without covariates. These tests commonly used for efficacy and treatment comparison outputs i.e. checking for efficacy of individual treatment or comparing multiple treatments.

The application is very interactive and flexible. It gives user the freedom to specify the statistical test or model of his/her choice (as per the specifications), set the desired significance level and specify the necessary parameters for selected test. This application accepts data in SAS, Excel and CSV format.

This application helps Statisticians to review the results, without doing independent programming. This will save time and will help Statisticians efficiently review the outputs.

In this paper we discuss use and features of this application with some illustrations.

WHY RSHINY?

R is an object-oriented programming language popular among Statisticians. While best known for graphs and visualization, it can admirably support performing various statistical analyses because of huge community support and early availability of new statistical techniques.

Shiny is a framework for building interactive applications using the R language. It's an R library installed on your local machine for creating interactive apps. It can be used to build useful web applications with only a few lines of code without knowledge of java. These applications are automatically "live" in the same way that spreadsheets are live. Outputs change instantly as users modify inputs, without requiring a reload of the browser.

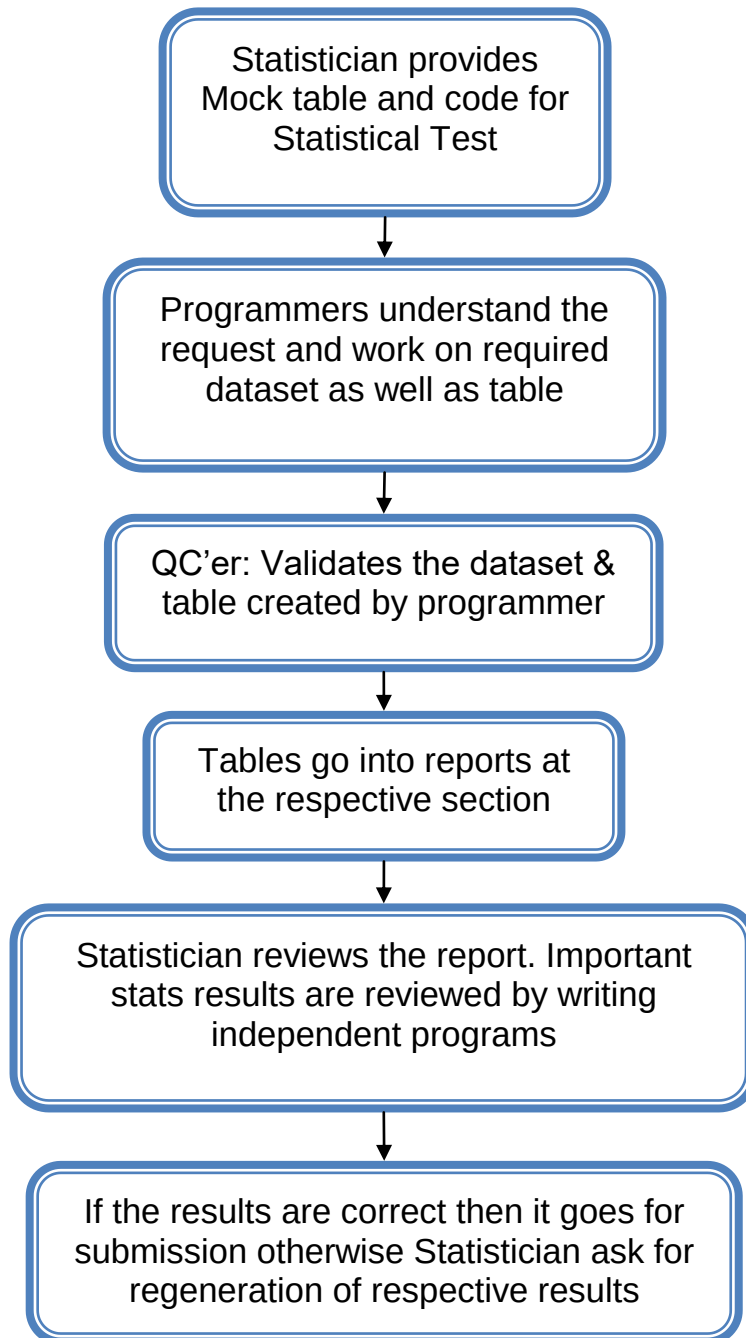
In this paper, we demonstrate the power of R together with Shiny to create an application that helps Statisticians to review critical statistical outputs. The application helps statistician to get the results within a few minutes and cross-

PhUSE EU Connect 2018

check the results in the report. This will help Statisticians perform their work efficiently and save time without compromising the quality.

PROCESS FLOW FOR STATISTICAL REVIEW

The general process flow for any report is represented in the below diagram-



PROCESS FLOW FOR STATISTICAL REVIEW USING APPLICATION

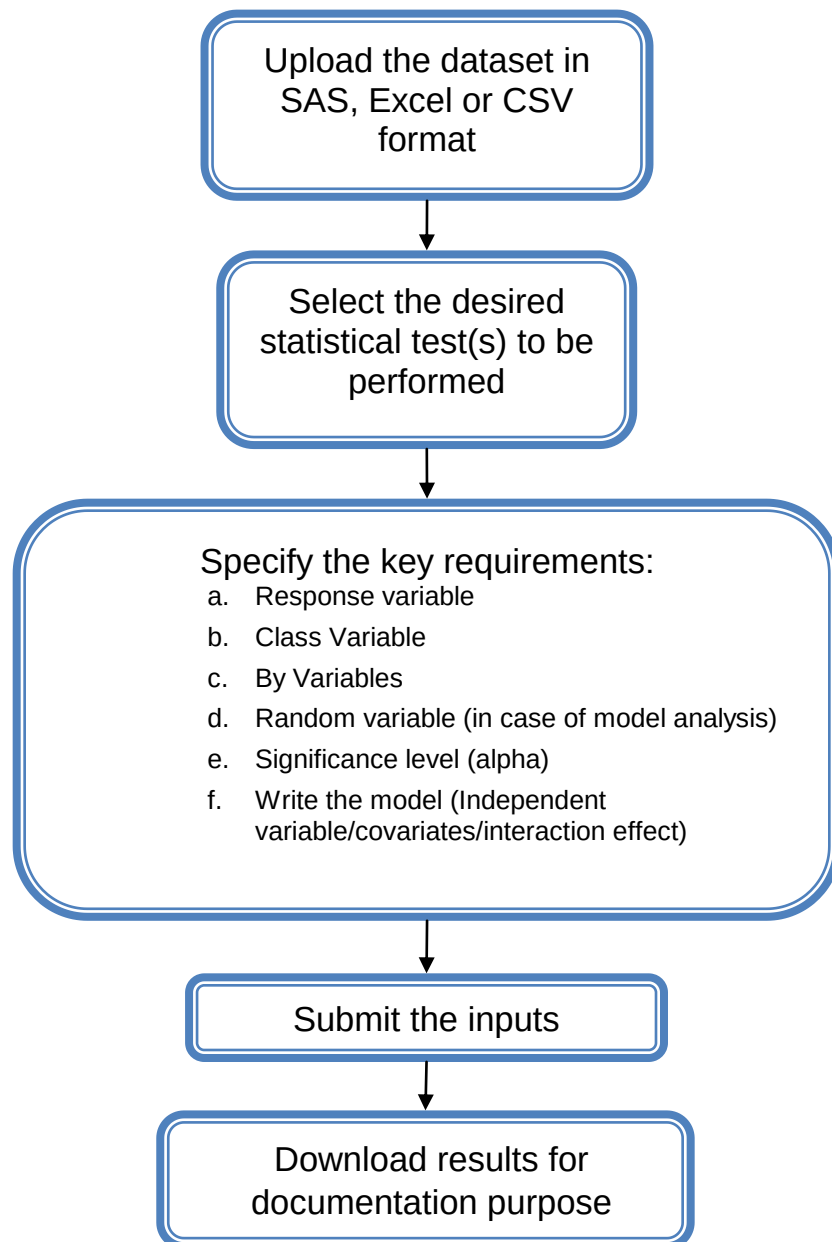
For the review step in the process discussed above, one can use this application instead of writing independent programming. This application is capable of performing a range of statistical tests; user can select the required ones

- One Sample t-Test
- Two Sample t-Test

PhUSE EU Connect 2018

- Paired t-Test
- Wilcoxon Signed Rank Pratt-Lehman Test
- Wilcoxon Signed Rank Test
- Mann-Whitney U test
- Friedman test
- PK Parameter Comparison
- Dose Proportionality

You have the flexibility to use SAS, Excel or csv dataset. Once dataset is finalized user can select the desired statistical test and provide the necessary information required to perform that test. Let's look at the steps required to be followed for use of this application –



PhUSE EU Connect 2018

After following the above process the results get generated and displayed in the application. You can cross-check the results with those in the reports and complete the review process.

Use of this application can improve the overall review process by reducing the time and efforts. In addition to this the review process gets strengthened by cross-checking the results using software (R) that is different from that used for production (which is SAS). For purpose of documentation related to review you can also download these results in the CSV format. We have deployed the use of this application for different statistical tests. Results corresponding to different tests have been discussed in further sections.

ILLUSTRATION 1

In this illustration, we are demonstrating comparison of two treatments A and B from a parallel arm design study. The endpoint of interest was Skin Appearance measured on continuous scale by a particular device. Here, as per the statistical analysis plan 'Two Sample t-test for unpaired data' was used. The hypothesis of interest was – Two treatments show different effects.

And now let's look at the results generated by programmers as below -

Treatment Comparison- Two-Sample t-test Skin Appearance: CFB Values						
Comparison	Timepoint	Mean Difference	t-Value	95% LCL	95% UCL	p-Value (Two Sided)
Treatment A Vs. Treatment B	15 Minutes Post Product application	-0.15	-1.01	-0.45	0.15	0.3150
	Day 4	-0.18	-0.18	-0.14	0.11	0.8553
	Week 1	-0.083	-1.28	-0.21	0.046	0.2048
	Week 4	-0.056	-1.14	-0.15	0.042	0.2575
	Week 8	-0.059	-0.73	-0.22	0.10	0.4675
	Week 12	-0.36	-3.05	-0.59	-0.12	0.0030

SE = Standard Error, LCL = Lower Confidence Limit, UCL = Upper Confidence Limit

As a reviewer author has generated the results using this RShiny application, refer to the screenshot below –

endpoint	timenum	timepoint	Product_pair	Mean	tValue	Probt	alpha	lower	upper
1 Skin Appearance	2	15 Minutes Post Product Application	Treatment A - Treatment B	-0.15032980	-1.0107607	0.315011498	0.05	-0.4460613	0.14540174
2 Skin Appearance	3	Day 4	Treatment A - Treatment B	-0.01137592	-0.1828323	0.855391332	0.05	-0.1351954	0.11244357
3 Skin Appearance	4	Week 1	Treatment A - Treatment B	-0.08298167	-1.2766700	0.204799610	0.05	-0.2120034	0.04604004
5 Skin Appearance	5	Week 4	Treatment A - Treatment B	-0.05632175	-1.1395597	0.257496817	0.05	-0.1545128	0.04186933
6 Skin Appearance	6	Week 8	Treatment A - Treatment B	-0.05923936	-0.7303770	0.467523177	0.05	-0.2209174	0.10243871
4 Skin Appearance	7	Week 12	Treatment A - Treatment B	-0.35908742	-3.0456152	0.003039947	0.05	-0.5932979	-0.12487689

Now we can clearly observe the mean difference, t statistic, p-value and 95% confidence interval matching in both these results. This shows that the results generated by programmers as well as by this application are the same, which indicates that the production results are correct.

Now, let's look at the snapshot below:

PhUSE EU Connect 2018

Two Sample T-Test

The TTEST Procedure

Variable: CFB

Endpoint=Skin Appearance Timenum=2 Timepoint=15 Minutes Post Product Application

Treatment	N	Mean	Std Dev	Std Err	Minimum	Maximum
Treatment A	50	1.1332	0.5895	0.0834	0.1130	3.1270
Treatment B	49	1.2835	0.8622	0.1232	0.1630	4.0270
Diff (1-2)		-0.1503	0.7371	0.1482		

Treatment	Method	Mean	95% CL Mean	Std Dev	95% CL Std Dev
Treatment A		1.1332	0.9656 1.3007	0.5895	0.4924 0.7346
Treatment B		1.2835	1.0358 1.5311	0.8622	0.7190 1.0771
Diff (1-2)	Pooled	-0.1503	-0.4444 0.1438	0.7371	0.6464 0.8577
Diff (1-2)	Satterthwaite	-0.1503	-0.4461 0.1454		

Method	Variances	DF	t Value	Pr > t
Pooled	Equal	97	-1.01	0.3129
Satterthwaite	Unequal	84.649	-1.01	0.3150

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	48	49	2.14	0.0091

This is a SAS output; observe that the p-value (0.0091) for testing equality of variances is significant. Hence, in this situation we need to consider the results of t-test from Satterthwaite method. Here, programmer should not consider the results from Pooled Method. One might say the p-values in both Pooled and Satterthwaite are approximately the same, but this might not be the case always. Especially, when p-value is near boundary line which is alpha (in this case 0.05). This concern is already addressed in RShiny application, which considers results from correct method.

ILLUSTRATION 2

Next illustration is for efficacy analysis of two treatments A and B from a parallel arm design study. The endpoint of interest was Skin Appearance Grade given by dermatologist on ordinal scale. Here, as per the statistical analysis plan 'Wilcoxon Signed Rank Pratt-Lehmann Test' was used. The hypothesis of interest was – Both the treatments are efficacious.

Now let's look at the results generated by programmers as below –

PhUSE EU Connect 2018

Efficacy Checking- Wilcoxon Signed Rank Pratt-Lehmann Test Skin Appearance Grade: CFB Values						
Treatment	Timepoint	N	Mean	SE	Z-Stat	p-Value (Two Sided)
Treatment A	15 Minutes Post Product application	49	-0.59	0.05	-6.23	<0.0001
	Day 4	49	-0.07	0.04	-1.94	0.0522
	Week 1	49	-0.09	0.05	-1.85	0.0639
	Week 4	48	-0.20	0.05	-3.43	0.0006
	Week 8	47	-0.18	0.06	-2.71	0.0066
	Week 12	47	-0.12	0.06	-1.87	0.0617
Treatment B	15 Minutes Post Product application	49	-0.58	0.04	-6.17	<0.0001
	Day 4	49	-0.21	0.05	-3.97	<0.0001
	Week 1	48	-0.17	0.05	-3.02	0.0025
	Week 4	47	-0.26	0.06	-3.52	0.0004
	Week 8	47	-0.39	0.07	-4.43	<0.0001
	Week 12	46	-0.43	0.08	-4.52	<0.0001

Z-stat and p-values are based on Wilcoxon Pratt Lehmann test and not on Mean and SE.
N= Number of Subjects, SE= Standard Error.

As a reviewer author has generated the results using this RShiny application, refer to the screenshot below –

	endpoint	treatment	timenum	timepoint	efficacy_stat	efficacy_pval
1	Skin Appearance Grade	Treatment A	2	15 Minutes Post Product Application	-6.227936	4.726197e-10
3	Skin Appearance Grade	Treatment A	3	Day 4	-1.941451	5.220364e-02
5	Skin Appearance Grade	Treatment A	4	Week 1	-1.853162	6.385917e-02
9	Skin Appearance Grade	Treatment A	5	Week 4	-3.429263	6.052220e-04
11	Skin Appearance Grade	Treatment A	6	Week 8	-2.713979	6.648037e-03
7	Skin Appearance Grade	Treatment A	7	Week 12	-1.868636	6.167352e-02
2	Skin Appearance Grade	Treatment B	2	15 Minutes Post Product Application	-6.170182	6.821142e-10
4	Skin Appearance Grade	Treatment B	3	Day 4	-3.969318	7.207878e-05
6	Skin Appearance Grade	Treatment B	4	Week 1	-3.019576	2.531291e-03
10	Skin Appearance Grade	Treatment B	5	Week 4	-3.515093	4.396003e-04
12	Skin Appearance Grade	Treatment B	6	Week 8	-4.428620	9.483775e-06
8	Skin Appearance Grade	Treatment B	7	Week 12	-4.520878	6.158380e-06

Now we can clearly observe the z-statistic (efficacy_stat) and p-value (efficacy_pval) matching in both these results. This shows that the results generated by programmers as well as by the application are same, which indicates that the production results are correct. The results generated by programmers also include N, Mean, and SE along with Z statistic and p-value. This table has a footnote mentioning that the results are based on Wilcoxon Pratt-Lehman test; not on mean and SE. However, our application is only providing Z statistic and p-value; which is of topmost importance.

WHY WILCOXON PRATT-LEHMAN TEST?

Wilcoxon Signed Rank Test excludes zeroes (if any) from test which results in decrease in effective sample size. To handle this situation the new version of this test has been proposed by Pratt-Lehmann. In this version there is a utility to adjust zeroes. And we were dealing with ordinal data in this case.

ILLUSTRATION 3

Here, we demonstrate comparison of two treatments A and B from a parallel arm design study. The endpoint of interest was Skin Appearance Grade given by dermatologist on ordinal scale. Here, as per the statistical analysis plan 'Mann-Whitney (Wilcoxon Rank Sum) Test' was used. The hypothesis to be tested was – Two treatments show different effects.

Below is the screenshot of output created as per the plan –

Treatment Comparison - Wilcoxon Rank Sum Test (Mann-Whitney Test)										
Skin Appearance Grade: CFB Values										
Comparison	Timepoint	N		Mean		Sum of Scores		Mean Difference (Treatment A – Treatment B)	Z-Stat	p-Value (Two Sided)
		Treatment A	Treatment B	Treatment A	Treatment B	Treatment A	Treatment B			
Treatment A Vs. Treatment B	15 Minutes Post Product application	49	49	-0.59	-0.58	2422.50	2428.50	-0.01	-0.02	0.9836
	Day 4	49	49	-0.07	-0.21	2708.00	2143.00	0.14	2.36	0.0183
	Week 1	48	49	-0.09	-0.17	2555.00	2198.00	0.07	1.28	0.1996
	Week 4	47	48	-0.20	-0.26	2365.00	2195.00	0.06	0.49	0.6253
	Week 8	47	47	-0.18	-0.39	2499.00	1966.00	0.21	2.13	0.0330
	Week 12	46	47	-0.12	-0.43	2554.50	1816.50	0.32	2.81	0.0050

Z-stat and p-values are based on Wilcoxon Rank Sum test and not on Mean.
N= Number of Subjects.

As a reviewer author has generated the results using this RShiny application, refer to the screenshot below –

endpoint	Timenum	timepoint	n1	n2	mean1	mean2	mean_diff	sum.rank.a1	sum.rank.b1	Z	p_value
1 Skin Appearance Grade	1	15 Minutes Post Product Application	49	49	-0.59183674	-0.5816327	-0.01020408	2422.5	2428.5	-0.02050425	0.983641126
2 Skin Appearance Grade	2	Day 4	49	49	-0.07142857	-0.2142857	0.14285714	2708.0	2143.0	2.36003369	0.018273276
3 Skin Appearance Grade	3	Week 1	49	48	-0.09183674	-0.1666667	0.07482993	2555.0	2198.0	1.28261613	0.199626595
4 Skin Appearance Grade	4	Week 4	48	47	-0.19791667	-0.2553191	0.05740248	2365.0	2195.0	0.48833401	0.625313277
5 Skin Appearance Grade	5	Week 8	47	47	-0.18085106	-0.3936170	0.21276596	2499.0	1966.0	2.13230804	0.032981532
6 Skin Appearance Grade	6	Week 12	47	46	-0.11702128	-0.4347826	0.31776133	2554.5	1816.5	2.80658282	0.005007004

What we observe here, is the output generated via RShiny application includes N, Mean, Sum of scores and Mean difference also. And values are matching in both the outputs. Statistics in the above screenshot i.e. n1, mean1, sum.rank.a1 refers to Treatment A; and n2, mean2, sum.rank.b1 refers to Treatment B whereas mean_diff refers to mean difference treatments A and B.

This also tells that we can customize our outputs as per need in this application.

ILLUSTRATION 4

Let's now move to the model-based analysis part. This application supports model-based analysis for some PK (pharmacokinetic) analysis like Summary and Comparison of PK parameters and Dose proportionality.

In this example, as per statistical analysis plan, model was natural log transformed PK parameter values as response variable, treatment as fixed effect, baseline as co-variate and subject as random effect at confidence level of 90%.

Reviewer needs to be very careful while cross-checking the results generated by the programmers in case of model-based outputs. This is because the results generated via software need to be back transformed for creating the required table. Back-transformation is required to get the results on the original scale which helps to correctly interpret the results.

This application takes care of this important aspect before displaying the results on the screen, hence it is very useful for the statisticians to diagnose if the mistake is due to back transformation.

PhUSE EU Connect 2018

Below is the screenshot of output created by programmers (Note: First footnote mentioning about the back transformation) -

Summary of Pharmacokinetic Parameters

Pharmacokinetic Parameter	Treatment					
	Treatment A			Treatment B		
	N	GM	90% CI	N	GM	90% CI
AUC _{inf} [†] (hr.ng/mL)	24	20.70	(19.07, 22.46)	24	24.82	(23.17, 26.59)
AUC _{Last} [†] (ng/mL)	23	20.41	(18.81, 22.14)	21	24.49	(22.83, 26.26)
C _{max} [†] (ng/mL)	24	2.75	(2.50, 3.03)	23	3.36	(3.13, 3.60)
[†] Back-transformed least squares mean and confidence interval from ANCOVA model performed on natural log-transformed values.						
GM: Geometric Mean CI: Confidence Interval						

Here is the snapshot of output created by our application -

	Param	gm_trta	lower_trta	upper_trta	gm_trtb	lower_trtb	upper_trtb
1	AUCInf	20.697841	19.072186	22.46206	24.825366	23.172783	26.595804
2	AUCLast	20.409305	18.809151	22.14559	24.488681	22.836083	26.260873
3	Cmax	2.749592	2.495637	3.02939	3.355265	3.129544	3.597266

We can observe that the highlighted results are matching in both the outputs.

ILLUSTRATION 5

The last illustration is about the Dose Proportionality assessment of the PK parameters. As per the statistical analysis plan Power Model was used to check for dose proportionality. The response variable was PK parameter results and independent variable was dose both on natural logarithmic scale.

Assessment of Dose Proportionality

Pharmacokinetic Parameter	Slope Estimate (90% CI)
AUC _{Last} [†] (nM·hr)	0.45 (0.16, 1.25)
CI= Confidence Interval	

Results coming from application are -

	X	slope	upper	lower
1	AUCLast	0.4508415	0.1627062	1.249234

Here again we can observe the slope and CI are matching in both the snapshots. The results obtained are back transformed, as model uses log scale data.

CONCLUSION

From all the illustrations discussed in the paper, we that this application can be used to review outputs from basic statistical tests like t-test, Mann-Whitney test etc. to the complex model based analysis used in PK analysis. This simply saves a lot of time, effort and gives the output in a click. This can be used to catch potential programming errors that may occur. This will not only ensure if the results are correct but also strengthen the review process as totally different software is used for the application.

Use of this application helps Statisticians to review the critical statistical outputs and save time, especially in the crunch timelines.

REFERENCES

1. Shiny R Studio Gallery
<https://shiny.rstudio.com/gallery/>
2. Shiny Documentation
<https://www.rdocumentation.org/packages/shiny/versions/1.1.0>
3. R Shiny Articles
<https://shiny.rstudio.com/articles/>
4. Application Development Framework for R/Shiny - PharmaSUG
<https://www.pharmasug.org/proceedings/2018/AD/PharmaSUG-2018-AD24.pdf>

ACKNOWLEDGMENTS

Thanks to all reviewers and our colleagues who provided there valuable time and feedback.

CONTACT INFORMATION

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

Author Name – Sameer Bamnote
Company – Cytel Statistical Software & Services Pvt Ltd
Address – SN 150, Lohia Jain IT Park, 6th Floor, A wing, Paud Road, Kothrud
City / Postcode – Pune, 411038
Work Phone: +91-20 - 67090146
Fax: +91-20-67090120
Email: sameer.bamnote@cytel.com
Web: www.cytel.com

Author Name – Adarsh Nagare
Company – Cytel Statistical Software & Services Pvt Ltd
Address – SN 150, Lohia Jain IT Park, 6th Floor, A wing, Paud Road, Kothrud
City / Postcode – Pune, 411038
Work Phone: +91-20 - 67090175
Fax: +91-20-67090120
Email: adarsh.nagare@cytel.com
Web: www.cytel.com

Brand and product names are trademarks of their respective companies.