

# Streamlining Study Design and Statistical Analysis

Ramkiran Gouripeddi<sup>1,2</sup>, Mollie Cummins<sup>1,2,3</sup>, Randy Madsen<sup>2</sup>, Bernie LaSalle<sup>1,2</sup>, Andrew Redd<sup>2,4</sup>, Xiangyang Ye<sup>1,2,5</sup>, Angela Presson<sup>2,4</sup>, Tom Greene<sup>2,6</sup>, Julio Facelli<sup>1,2</sup>  
<sup>1</sup>Department of Biomedical Informatics, <sup>2</sup>Center for Clinical and Translational Science, <sup>3</sup>College of Nursing; <sup>4</sup>Department of Internal Medicine, <sup>5</sup>Pharmacotherapy Outcome Research Center, <sup>6</sup>Department of Population Health Science, University of Utah, Salt Lake City, Utah, USA

## Introduction

- Key factors causing irreproducibility of research include those related to inappropriate study design methodologies and statistical analysis<sup>1</sup>.
- In modern statistical practice irreproducibility could arise due to:
  - Statistical: False discoveries, p-Hacking, Overuse/misuse of p-values, Low power, Poor experimental design.
  - Computational: Data, Code & Software management issues<sup>2</sup>.
- These require understanding the processes and workflows practiced by an organization, and the development and use of metrics to quantify reproducibility.

## Conclusion

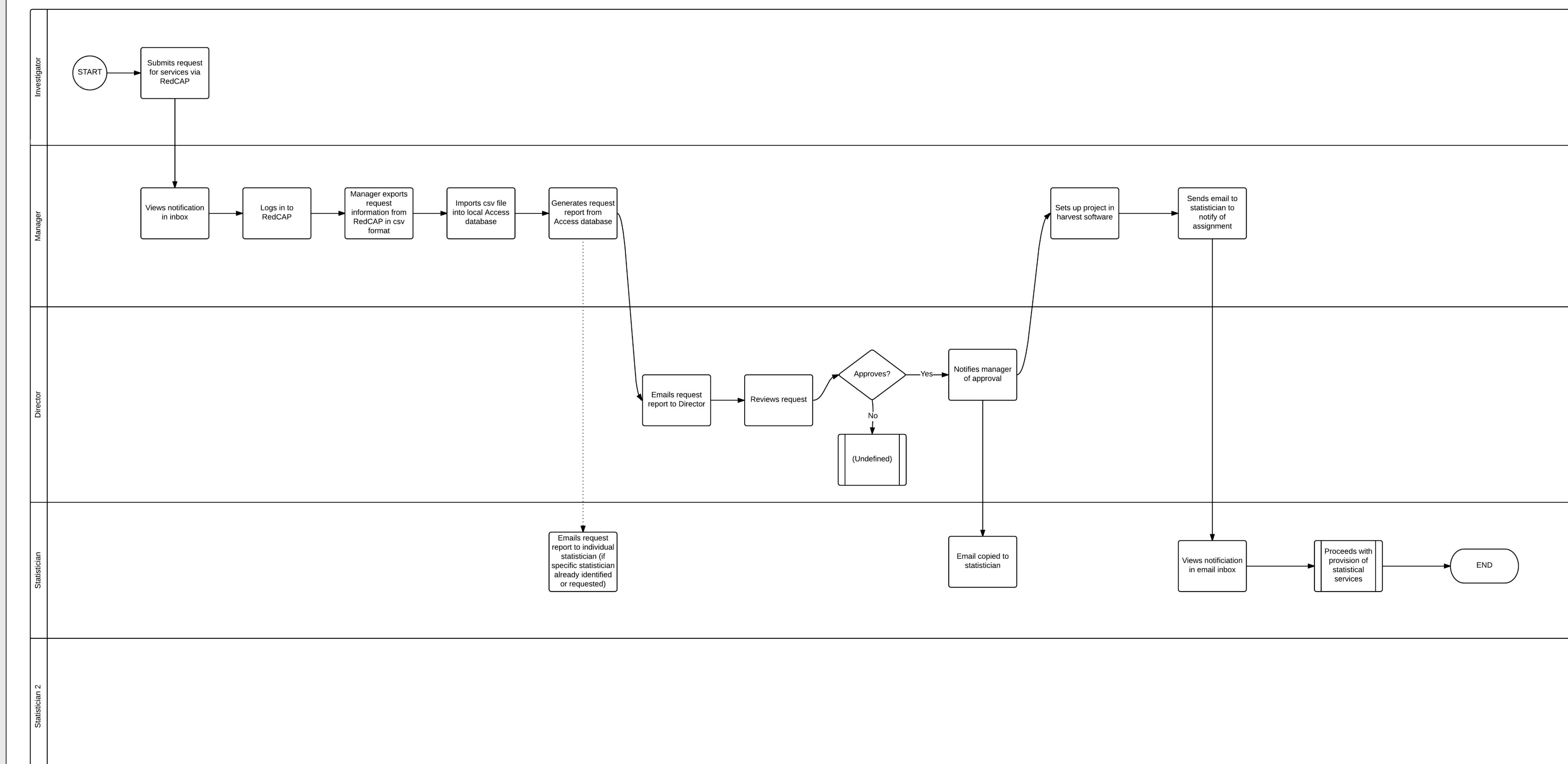
- Automated platforms for statistical workflows.
- Data-intensive process use process-workflow management platforms: Activiti<sup>3</sup>, Pegasus<sup>4</sup> and Taverna<sup>6</sup>.
- These strategies for sharing and publishing study protocols, data, code and results across the spectrum<sup>7</sup>, active collaboration with the research team, automation of key steps, along with decision support will ensure quality of statistical methods and reproducibility of research.

## Acknowledgements

This project is supported by NCATS UL1TR001067.


**Contact:** Ram Gouripeddi  
ram.gouripeddi@utah.edu

## Initiation of Statistical Services




## Example Projects, Sub-tasks Statuses and Tags for Measuring Reproducibility


Issue Types	Sub-tasks	Statuses	Tags
Default			
One-off Consultation	1. Consultation	1. Open 2. Closed	
Regular Project	1. Consultation 2. Write Statistical Analysis Plan (SAP) + QC Checks include Code review 3. Assign second consultant to reviews analysis plan for completeness, and appropriateness of methods 4. Approval of SAP by PI 5. Perform data analysis as outlined in SAP 6. Assign second consultant to perform data analysis as outlined in SAP 7. Check results from both analyses match 8. Submit results to PI 9. Approval of analysis by PI 10. Support for publication preparation 11. PI sends material for review 12. Review of publication	1. Open 2. Hold for PI 3. Data, access, dictionary received 4. Closed	1. Funded or not 2. Impact: High, medium, low

 Data Sharing
 


- Box

 Code Dissemination
 

- Bitbucket
- GitHub

 Documentation
 

- Confluence

 Workflow Tracking
 

- Jira

## Automation Platforms for Statistical Workflows

## JIRA for Project Tracking

## Methods

Within the Foundation of Discovery - Population Health Research, Center for Clinical and Translational Science, University of Utah, we are undertaking a project to streamline study design and statistical analysis workflows and processes.

- Elicited example statistical projects: We met with key stakeholders to understand the current practices.
- Understood organization structure and practices: We discussed these with the Foundation's leadership and Standards Committee.
- Defined key measurement points for assessing reproducibility: Study design, analysis plan, final report, requirements for quality checks, and double coding.
- Developed process *information* models: For different types of statistical needs using Lucidchart and Dia.
- Reviewed models for ideal workflows.
- Embed analytical and implement infrastructural approaches within the statisticians' workflows based on our finding.

## References

- Reproducibility and reliability of biomedical research', organised by the Academy of Medical Sciences, BBSRC, MRC and Wellcome Trust in April 2015.
- National Academies of Sciences, Engineering, and Medicine. Statistical Challenges in Assessing and Fostering the Reproducibility of Scientific Results: Summary of a Workshop. Washington, DC: The National Academies Press, 2016. doi:10.17226/21915.
- Aciviti, <http://activiti.org>
- Pegasus, <https://pegasus.isi.edu/>
- Apache Taverna. <https://taverna.incubator.apache.org/>
- Peng, R. 2011. Reproducible research in computational science. Science 334(6060):1226-1227