Assessing the impact of programming workshops on biomedical research reproducibility

Ariel Deardorff, Data Services Librarian
University of California, San Francisco Library

UCSF Library

Problem Statement

UCSF Library has taught 11 Software Carpentry programming workshops to over 700 researchers since 2016.

The workshops are popular but we do not know whether they impact researcher workflows or contribute to research reproducibility.

About the Workshops

2-day introductory programming workshops for researchers, featuring:

- R or Python
- Version control with Git
- Scripting with Unix

Based on Software Carpentry's open curriculum

Research Questions

1. Do introductory programming workshops lead to adoption of more computationally reproducible workflows in biomedical researchers?

2. Are there common pain points or steps in the research workflow that researchers want to improve through programming?

Methods

Semi-structured interviews with 14 UCSF workshop attendees before and (3 months) after participating in a Software Carpentry workshop.

Qualitative: Participants drew and described their research workflows (noting tools and methods) and discussed their goals for the workshop and their progress implementing workshop skills and tools.

Quantitative: Participants filled out a checklist of reproducible behaviors (publishing code, using version control, etc) to get a score out of 6.

Analysis 1: Why do researchers learn to program?

An initial thematic analysis of the qualitative data from the pre-workshop interviews. Found four major themes:

- Independence in data analysis
- Programming literacy
- New kinds of big data research
- Flexibility in tool choice

Published in the *Journal of the Medical Library Association*

Analysis 2: What is the impact of the workshops on researcher workflows?

Descriptive statistics and T-test to look for significant change in checklist scores

Thematic analysis to answer:

- What changes did researchers implement after the workshop?
- What enabled or prevented these changes?
- How has this impacted the reproducibility of their work?

Forthcoming article in *PLOS One*

Results - Quantitative

- Small but **not significant** increases in checklist scores (avg score increased from 1.6 to 2.2 out of 6)
- Takeaway: 3 months is not enough time to see complete overhaul in tools/methods used

Question	PreTest Total (n=14)	PostTest Total (n=12)
Use programming languages like R, Python, or the command line for data acquisition, processing, or analysis	7	8
Transform step-by-step workflows into scripts or functions	1	2
Use version control to manage code	0	2
Use open source software	7	10
Share your code publicly	5	2
Share your computational workflow or protocols publicly	3	2

Table 1. Checklist scores before and after the workshop

Results - Qualitative

New practices adopted/planned:

- Switching their data analysis/visualization tools to R/Python
- Using the command line for data exploration
- Exploring Github for collaboration
- Taking ownership/co-ownership of data analysis

Researcher talking about a new approach to working with their bioinformationist:

> “So, I think, I don't know if I'll be doing all of it, but at least more together. Really, I was handing everything off to her. She'd do everything on her computer and then I'd only see figures weeks later. And so **this would actually be like handling the data myself**, doing some in R, if I can't, or am having issues like helping her, helping me kind of troubleshoot those things. Or even if she eventually does do some of the analysis, I'll know what she's done specifically.”

Results - Qualitative

Other themes:

- Increase in programming literacy
- Interest in future learning

Researcher describing feeling more comfortable talking about programming with their colleagues:

> “I could understand a little bit more what the more informatics people in my lab are doing day to day and then they talk about stuff and I'm like ‘**Oh, I know those words’** like you sort of get to know the techniques that they're using a little bit more and the different software and stuff “

Results - Qualitative

Factors that promoted new skills:

- Skilled collaborators/colleagues
- Immediate research need

Barriers to implementing new skills:

- PI/collaborators resistance to new tools
- Lack of time

Researcher talking about lack of time to implement what they learned:

> “So after, right after attending the workshop, I started using Python. But the thing is obviously I had to fix some problems. So, because I had the, I knew the basics, but then when I needed to do my own things and ask specific questions, I was not able to. **And so, in order to be quick and to get my things done without wasting too much time …. I just continued with my usual way of proceeding.”**

Conclusion

While none of the researchers completely changed their workflows, all of them learned or tried something that would make their research more open source, transparent, and reproducible.

Introductory programming workshops can be an excellent way for libraries and other organizations to contribute to biomedical research reproducibility.

Libraries should focus on follow-up workshops and continued engagement to keep researchers motivated to implement new practices

Learn More

Deardorff, A. (forthcoming). Assessing the impact of introductory programming workshops on the computational reproducibility of biomedical workflows. *PLOS One*.

Deardorff, A. (2020). Why do Biomedical Researchers Learn to Program? An Exploratory Investigation. *Journal of the Medical Library Association. 108*(1), 29. DOI: 10.5195/jmla.2020.819

Research Data Available in Dryad

Impact of Programming Workshops on Biomedical Reproducibility (IPWBR)

Deardorff, Ariel, University of California, San Francisco, https://orcid.org/0000-0001-8930-6089
Ariel.Deardorff@ucsf.edu
Publication date: March 17, 2020
Publisher: UC San Francisco
https://doi.org/10.7272/Q6RV0KW6

Citation

Deardorff, Ariel (2020), Impact of Programming Workshops on Biomedical Reproducibility (IPWBR), v6, UC San Francisco, Dataset, https://doi.org/10.7272/Q6RV0KW6

Abstract

The objective of this study was to assess the impact of introductory programming workshops on the computational reproducibility of biomedical research workflows. This mixed methods study consisted of in-depth interviews with 14 biomedical researchers before and after participation in an introductory programming workshop. During the interviews, participants descreir research workflows and responded to a quantitative checklist measuring reproducible behaviors. The interview data was analyzed using a thematic analysis approach, and the pre and post workshop checklist scores were compared to assess the impact of the workshop on computational reproducibility of the researchers' workflows.

Download dataset ~ 4 MB

Download Data Publication (PDF)

Data Files

- December 11, 2019
- March 17, 2020

Metrics

58 views

6 downloads

0 citations

Keywords

biomedical research

Thanks to:

- Kristine R. Brancolini, Marie Kennedy, and the Institute for Research Design in Librarianship (IRDL)
- Savannah Kelly, Jill Barr-Walker, and Catherine Nancarrow for feedback and support

Questions?

Email me! Ariel.deardorff@ucsf.edu