Open Sciences Survey-Email

Start of Block: Questions

Q1 FOR IRB USE ONLY  
 IRB ID #: 201703045  
 APPROVAL DATE: 07/03/17  
 RELEASED DATE: 07/03/17  
 EXPIRATION DATE: N/A   
   
   
   
   
We invite you to participate in a research study being conducted by investigators from Washington University in St. Louis. The purpose of the study is to gain a better understanding of how people work with quantitative data in public health in order to develop recommendations to increase the use of reproducible research practices.

If you agree to participate, we would like you to take a brief survey asking about your data management, analysis, and organizational procedures. You are free to skip any questions that you prefer not to answer. It will take approximately 15 minutes to complete the survey.

This survey is completely anonymous. We will not collect your name or any identifying information about you. It will not be possible to link you to your responses on the survey.   
Once you complete the survey, you will be eligible to be entered into a drawing for a $100 Amazon.com gift card. You must complete the entire survey to be eligible. You will be directed to a separate unlinked survey to collect your name and contact information for the drawing. We hope to conduct the drawing within the next two months.

Taking part in this research study is completely voluntary. You may choose not to take part at all. If you decide to be in this study, you may stop participating at any time. Any data that were collected as part of your participation in the study will remain as part of the study records and cannot be removed.

If you do not wish to participate in this study or want to end your participation in the study, please click “I do not wish to participate” at the bottom of the page and you will be exited out of the survey. You will not be penalized or lose any benefits for which you otherwise qualify.   
We encourage you to ask questions. If you have any questions about the research study itself, please contact Bobbi Carothers at bcarothers@wustl.edu or 314-935-3745. If you have questions, concerns, or complaints about your rights as a research participant, please contact the Human Research Protection Office at 660 South Euclid Avenue, Campus Box 8089, St. Louis, MO  63110, 1-(800)-438-0445 or email hrpo@wustl.edu. General information about being a research participant can be found on the Human Research Protection Office website, http://hrpo.wustl.edu/. To offer input about your experiences as a research participant or to speak to someone other than the research staff, call the Human Research Protection Office at the number above.

Thank you very much for your consideration of this research study.

* I have read this consent letter and agree to participate (1)
* I do not wish to participate (2)

Skip To: End of Block If Q1 = [2]

Q2 Research Reproducibility in Public Health

Q3 This survey is part of an open science project to better understand what strategies public health professionals use that may contribute to research reproducibility. We are interested in:

data management, code annotation and organization, manuscript writing practices, sharing of data and statistical code following manuscript publication, and other factors that encourage reproducible research. The results of the survey will be used to inform the development of recommendations and training for journals and scientists to increase the use of reproducible research practices. You can follow us or tweet to us using @coding2share.

Q4 Job Responsibilities   
What are your primary job responsibilities related to quantitative data? *(Check all that apply)*

* Data collection (1)
* Data management (2)
* Descriptive analysis (3)
* Inferential analysis (4)
* Visualization/graph production (5)
* Contributing to publications or reports (6)
* ⊗I do not work with quantitative data in my job (7)

Skip To: End of Block If Q4 = [7]

Display This Question: If Q4 != [6]

Q5 Have you recently worked on any publications or reports where you were the primary analyst or held another role involved in the development or execution of statistical code?

* Yes (1)
* No (2)

Skip To: End of Block If Q5 = [2]

Q6 Many questions in this survey will ask you to describe your work on a specific recent publication or report where you led, or were involved in, statistical analyses. Please identify a specific publication or report where you led, or were involved in, statistical analyses. Provide a nickname for this publication or report to help facilitate your recall during the survey (e.g. "AJPH obesity paper"). The nickname will appear in questions during this survey but will not be used in any other way. Nickname for a recent publication or report:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

|  |
| --- |
|  |

Q7 Data and code file management practices   
While working on your recent publication or report, how did you store the data and statistical code used for analyses? *(Check all that apply)*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | In the cloud (e.g., Dropbox, GitHub) (1) | On a local server at your workplace (2) | On a desktop computer (3) | On a laptop computer (4) | On a removable storage device (e.g., USB drive, CD) (5) | There were no statistical code files (e.g., used point-and-click/menu driven approach in SPSS, Excel, or other software) (6) |
| Data (1) |  |  |  |  |  |  |
| Statistical code (2) |  |  |  |  |  |  |

Q8 While working on your recent publication or report, where did you store the files? *(Check all that apply)*

|  |  |  |
| --- | --- | --- |
|  | In project-specific folders or directories (1) | In folders or directories not specific to the project (2) |
| Data (1) |  |  |
| Q7 != Statistical code [6]  Statistical code (2) |  |  |

Q9 While working on your recent publication or report, how did you organize your files?

|  |  |  |
| --- | --- | --- |
|  | In a single file (1) | In multiple files (2) |
| Data (1) |  |  |
| Q7 != Statistical code [6]  Statistical code (2) |  |  |

Q10 While working on your recent publication or report, how would you describe file accessibility? *(Check all that apply)*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Shareable with anyone (public) (1) | Shareable but secure (i.e., authorized access only) (2) | Private (3) |
| Data (1) |  |  |  |
| Q7 != Statistical code [6 ]  Statistical code (2) |  |  |  |

Q11 If you were to leave your institution suddenly, how easy would it be for someone to find the specific files for your recent publication or report without any instruction from you?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Very easy (1) | Somewhat easy (2) | Somewhat difficult (3) | Very difficult (4) |
| Data (1) |  |  |  |  |
| Q7 != Statistical code [6]  Statistical code (2) |  |  |  |  |

Q12 Which of the following best describes how and where your store your data and statistical code **in general** (not specific to your recent publication or report)?

|  |  |  |  |
| --- | --- | --- | --- |
|  | No set process (1) | A set process developed in-house (2) | A set process developed externally (3) |
| Data (1) |  |  |  |
| Statistical code (2) |  |  |  |

Display This Question: If Q7 != Statistical code [6]

Q13 Code annotation and documentation

The next set of questions ask about practices related to annotation of your code file for your recent publication or report.

Q14 A variable dictionary, or codebook, lists the variables in a data set and how they were measured and stored. Is there a variable dictionary for the data used for your recent publication or report?

* Yes (1)
* No (2)

Display This Question: If Q7 != Statistical code [6]

Q15 A prolog is a block of comments at the top of a code file that includes general information about the file. It can also be called a header, preface, or introduction. Did you include a prolog at the beginning of the statistical code used for your recent publication or report?

* Yes (1)
* No (2)

Display This Question: If Q15 = [1]

Q16 Which of the following were included in the prolog for your recent publication or report? *(Check all that apply)*

* Project name (1)
* Code developer name (2)
* Contact information for the code developer or another author (3)
* The purpose of the code (4)
* Information about the language/program used to run the code (e.g., software package, version of language) (5)
* File paths or other locating information for the data (6)
* File paths or other locating information for the statistical code (7)
* The last date the code was edited (8)
* Something else (please specify): (9)

-------------------------------------------------------------------------------------------------------------------------------

Display This Question: If Q7 != Statistical code [6]

Q17 Did you include comments in your statistical code for your recent publication or report?

* Yes (1)
* No (2)

Skip To: Q22 If Q17 = [2]

Display This Question: If Q7 != Statistical code [6]

Q18 Which of the following is the closest to describing your comments in the code for your recent publication or report?

* *Comprehensive*: most or all steps of analysis have comments (1)
* *Moderate*: comments used regularly throughout the code for new sections and clarification (2)
* *Sparse*: few or no comments used (3)

Display This Question: If Q7 != Statistical code [6]

Q19 Did you follow any specific guidelines, recommendations, or style guides during statistical code development for your recent publication or report? (e.g., Google’s R style guide; SAS style guide)

* Yes, closely followed guidelines (1)
* Yes, partially followed guidelines (2)
* No (3)

Skip To: Q22 If Q19 = [3]

Display This Question: If Q7 != Statistical code [6]

Q20 Which guidelines did you use for your recent publication or report? *(Check all that apply)*

* Google’s R Style Guide (1)
* Bioconductor’s Coding Standards (R) (2)
* Hadley Wickham’s Style Guide (R) (3)
* Henrik Bengtsson’s R Coding Conventions (4)
* Colin Gillespie’s R Style Guide (5)
* SAS Style Guide (6)
* Guidelines for Coding of SAS Programs (7)
* Suggestions on Stata Programming Style (8)
* Google Python Style Guide (9)
* GNU Mailman Style Guide (Python) (10)
* Style Guide for Python Code: PEP 8 (11)
* Visual Basic Coding Conventions (Excel, Access) (12)
* MATLAB Style Guidelines (13)
* Something else (please specify): (14) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Display This Question: If Q7 != Statistical code [6]

Q21 What were the reasons for you following those guidelines for your recent publication or report? *(Check all that apply)*

* Policy of my research team/lab (1)
* Required for publication (2)
* Makes my life easier (3)
* It produces better code (4)
* I was taught this way (5)
* Improves collaboration (6)
* Increases reproducibility (7)
* Something else (please specify): (8) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Display This Question: If Q7 != Statistical code [6]

Q22 Did you use any of the following in organizing or formatting your statistical code for your recent publication or report? *(Check all that apply)*

* Used nouns for variables and/or verbs for functions (1)
* Limited lines of code to a certain length (e.g., 80 characters maximum) (2)
* Included metadata, such as the date or project name, in the file title (3)
* Included seed values for analyses that included randomness (4)
* Used a consistent way to name variables and functions (e.g., all lower case, camel case, meaningful words, etc) (5)
* Separated analysis steps with white space or blank lines (6)
* Used indentation to group lines of code within procedures/functions (7)
* Wrote functions for tasks repeated multiple times (8)
* Included some results within the annotation (e.g., mean age = 39, outcome did not meet normality assumption) (9)
* Integrated code with text and results into a single output, also known as literate programming (e.g., Sweave, Jupyter Notebook, R Markdown) (10)
* Something else (please specify): (11) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
* ⊗None of the above (12)

Q23 Reproducibility of statistical analyses   
The next set of questions ask about practices related to the reproducibility of statistical analyses based on the main text or supplemental files of your recent publication or report.

Q24 Which of the following were included in your recent publication or report?  *(Check all that apply)*

* Software used (1)
* Units of analysis (e.g., people, health departments) (2)
* Details on missing data handling (3)
* Variable recoding details (4)
* The name, units, and types of variable analyzed (5)
* The statistical approaches used (e.g., logistic regression, mixed-effects regression) (6)
* The type of test statistics computed (e.g., chi-squared, F) (7)
* The value of test statistics (8)
* The specific variables included in each statistical model (9)
* Sample sizes for each analysis (10)
* Precise p-values when possible (e.g., p=.02 rather than p (11)
* ⊗None of the above (12)

Q25 Following acceptance of your recent publication or report, which of the following did you create and which of the following did you make publicly available? *(Check all that apply)*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | A clean version of data used for analyses (1) | Q7 != Statistical code [ 6]  Clean statistical code for results included in the publication (2) | A project directory with data and/or statistical code used for the publication (3) | A readme file explaining the data and/or statistical code (4) | None of the above (5) |
| Created (1) |  |  |  |  |  |
| Made publicly available (2) |  |  |  |  |  |

Display This Question: If Q25 != Made publicly available [2]

Q26 Have you *ever* made your statistical code publicly available?

* Yes (1)
* No (2)

Display This Question: If Q26 = [2]

Q27 Have you published papers (**in general**, not specific to your recent publication or report) with publicly available data (either made public by yourself or public use data) over the last year?

* Yes (1)
* No (2)

Display This Question: If Q25 != Made publicly available [5]

Q28 Were you required to make your files of your recent publication or report publicly available? *(Check all that apply)*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Yes, by a funder (1) | Yes, by the journal (2) | Yes, by my employer (3) | Yes, by my research team (4) | No, it was not required (5) |
| Data (1) |  |  |  |  |  |
| Q7 != Statistical code [6]  Statistical code (2) |  |  |  |  |  |

Display This Question: If Q7 != Statistical code [6]

Q29 Which of the following best describes how the statistical code was developed to get the results for your recent publication or report?

* The code was developed by one person working alone (1)
* The code was developed by one person and other people checked the code (co-pilot strategy) (2)
* Code was developed in separate files by two or more people independently and then compared before choosing one or a combination of the two (i.e., parallel code development) (3)
* Code was developed by two or more people working together on the same coding file (4)

Q30 Research reproducibility facilitators and barriers   
The next set of questions about perceived barriers to reproducibility of published statistical analyses **in general** (not specific to your recent publication or report).

|  |
| --- |
|  |

Q31 In your opinion, which of the following are**most** important to facilitate increased use of reproducible research practices? *(Pick your top 2)*

* Additional human and financial resources (1)
* Training on reproducible research practices (2)
* Requirements by funders to disseminate data and statistical code (3)
* Requirements by journals to include access to data and statistical code (4)
* Workplace incentives (e.g., pay increases or more credit toward tenure for reproducible/reproduced work) (5)
* Something else (please specify): (6)

Q32 Which of the following barriers have you experienced when making data or statistical code publicly available **in general**? *(Check all that apply)*

* Lack of time (1)
* Lack of knowledge/training on reproducible research practices (2)
* Data privacy (3)
* Intellectual property concerns (4)
* Professional competition (5)
* Concerns of errors being discovered (6)
* Lack of incentive (7)
* Something else (please specify): (8)
* ⊗No barriers experienced (9)
* ⊗I have not tried to make data or statistical code available (10)

Q33 If training were available on reproducible research practices, how likely are you to participate?

* Very likely (1)
* Somewhat likely (2)
* Somewhat unlikely (3)
* Very unlikely (4)

Display This Question: If Q33 = [1] Or [2]

|  |
| --- |
|  |

Q34 What format would you prefer for a reproducible research training? *(Pick your top 2)*

* In-person workshop or short course (1)
* In-person one-on-one assistance (2)
* Online asynchronous course (3)
* Online live webinar (4)
* Documents (e.g., manual, collection of resources, checklists, etc) (5)
* Blog or other living web presence (6)
* Social media site (7)
* Something else (please specify): (8)

Q35 Participant characteristics    
The last set of questions asks general questions about your characteristics.

Q36 Identify the statistical software you use the most often:

* Excel (1)
* M-Plus (2)
* MatLab (3)
* Python (4)
* R (5)
* SAS (6)
* SPSS (7)
* Stata (8)
* Other (please specify) (9)

|  |
| --- |
|  |

Q37 What are the resources you prefer to use when you need to do something new in your statistical software? *(Pick your top 2)*

* Software user manual(s) (1)
* Textbook or reference book (2)
* Internet search (i.e. Google) (3)
* YouTube video demonstration or something similar (4)
* In-person one-on-one help (5)
* Email help (6)
* Help via a chat feature (gchat or similar) (7)
* Online training or course (8)
* In-person training or course (9)
* Websites, blogs (i.e. Stack Exchange) (10)
* Something else (specify): (11)

Q38 What is your highest earned degree?

* PhD (1)
* MD (2)
* MD/PhD (3)
* DrPH (4)
* DSc (5)
* EdD (6)
* MPH (7)
* MSW (8)
* MS or MSc (9)
* MRes or MPhil (10)
* MA (11)
* MBA (12)
* MAT (13)
* BS (14)
* BA (15)
* Something else (please specify): (16)

Q39 What is the field of your highest earned degree?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Q40 What is your job title?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Q41 How long have you been in your current position?

* Less than 1 year (1)
* 1-3 years (2)
* 4-10 years (3)
* More than 10 years (4)

Q42 What type of organization are you in?

* University/school (1)
* For-profit (2)
* Government (3)
* Nonprofit (4)
* Something else (please specify): (5)

Q43 How do you describe yourself?

* Male (1)
* Female (2)
* Transgender (3)
* Do not identify as female, male, or transgender (4)

Q44 Which categories describe you? *(Check all that apply)*

* White (1)
* Hispanic, Latino, or Spanish origin (2)
* Black or African American (3)
* Asian (4)
* American Indian or Alaska Native (5)
* Middle Eastern or North African (6)
* Native Hawaiian or Other Pacific Islander (7)
* Some other race, ethnicity, or origin (8)

|  |
| --- |
|  |

Q45 Please enter your age in years:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Q46 Do you have any suggestions you'd like to share with us about the reproducible research?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

End of Block: Questions

Start of Block: Thanks message

Q47 Thank you for your participation. Please click "submit" below to submit your responses. Feel free to contact us at coding2share@gmail.com with questions or feedback and follow us or tweet to us using @coding2share.

End of Block: Thanks message