

Guide for Accelerating Computational Reproducibility in the Social Sciences



BERKELEY INITIATIVE FOR TRANSPARENCY
IN THE SOCIAL SCIENCES

ACRE Team

2022-09-20

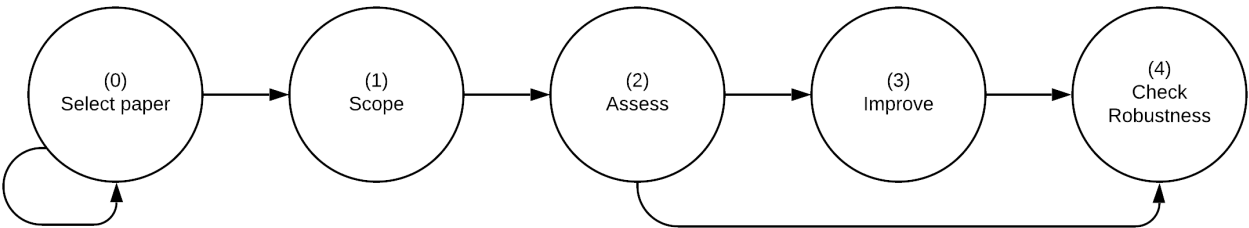
Contents

	3
Introduction	4
Beyond binary judgments	4
Reproduction stages	6
Reproduction strategies	7
1 Selecting a paper	8
1.1 From candidate to declared paper	8
1.2 Identifying the relevant timeline	9
1.3 Potential sources of papers to prioritize	9
2 Scoping	13
2.1 Read and summarize the paper	14
2.2 Record a revised reproduction package	14
2.3 Record scope of the exercise	15
3 Assessment	17
3.1 Describe the inputs	17
3.2 Connect display items to all its inputs usign the Diagram Builder	20
3.3 Assign a reproducibility score.	22
4 Improvements	26
4.1 Display item improvements	26
4.2 Paper-level improvements	28
4.3 Documenting the improvements using version control	29
5 Checking for Robustness	30
5.1 Feasible robustness checks: increasing the number of feasible specifications	32
5.2 Justifying and testing reasonable robustness checks	33

<i>CONTENTS</i>	3
6 Concluding the Reproduction	34
6.1 Outputs	34
6.2 Visibility and data use	35
7 Guidance for Instructors Supervising Reproduction Assignments	36
7.1 Common use cases	37
7.2 Paper selection and assignment timelines	37
7.3 Grading strategy	38
7.4 Facilitating <i>class anonymous</i> reproductions	40
8 Guidance for Constructive Communication Between Reproducers and Original Authors	41
8.1 For reproducers contacting the authors of the original study	42
8.2 For original authors responding to requests from reproducers	47
8.3 Harassment and/or discrimination	48
9 Examples of Reproduction Trees	50
9.1 Stylized examples	50
9.2 Examples from real reproduction attempts	52
10 Contributions	55
10.1 Contributing feedback on this Guide	55
10.2 List of Contributors: Guidelines content and source code:	55
10.3 Suggested citation format	56
10.4 Acknowledgments	56
11 Definitions	57

See a full list of contributors

A companion to the **Social Science Reproduction Platform**.



(0) Select	(1) Scoping	(2) Assessment	(3) Improvement		(4) Robustness
			Display-Item -Level	Paper-Level	
<ul style="list-style-type: none"> • Select paper 	<ul style="list-style-type: none"> • Read paper 	<ul style="list-style-type: none"> • Describe inputs 	<ul style="list-style-type: none"> • Raw data 	<ul style="list-style-type: none"> • Version control 	<ul style="list-style-type: none"> • Analytical Choices
<ul style="list-style-type: none"> • Search materials 	<ul style="list-style-type: none"> • Identify claims 	<ul style="list-style-type: none"> • Build reproduction diagrams 	<ul style="list-style-type: none"> • Analysis data 	<ul style="list-style-type: none"> • Documentation 	<ul style="list-style-type: none"> • Type of choices
<ul style="list-style-type: none"> • Check SSRP 	<ul style="list-style-type: none"> • Declare estimates 	<ul style="list-style-type: none"> • Assign a reproducibility score 	<ul style="list-style-type: none"> • Analysis code 	<ul style="list-style-type: none"> • Dynamic control 	<ul style="list-style-type: none"> • Choice value
<ul style="list-style-type: none"> • Declare/ Discard Paper 			<ul style="list-style-type: none"> • Cleaning code 	<ul style="list-style-type: none"> • File structure 	<ul style="list-style-type: none"> • Justify and test alternatives
			<ul style="list-style-type: none"> • Debug analysis code 		
			<ul style="list-style-type: none"> • Debug cleaning code 		

This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

Introduction

Computational reproducibility is the degree to which it is possible to obtain consistent results using the same input data, computational methods, and conditions of analysis (National Academies of Sciences 2019). In 2019, the American Economic Association updated its Data and Code Availability Policy to require that the AEA Data Editor verify the reproducibility of all papers before they are accepted by an AEA journal. Similar policies have been adopted in political science, particularly at the *American Journal of Political Science*. In addition to the requirements laid out in such policies, the data editors of several social science journals produced detailed recommendations and resources to facilitate compliance. The goal of such policy changes is to improve the computational reproducibility of *all* published research going forward, after several studies showed that rates of computational reproducibility in the social sciences range from somewhat low to alarmingly low (Galiani, Gertler, and Romero 2018; Chang and Li 2015; King et al. 2018).

This Guide includes a common approach, terminology, and standards for conducting *reproductions*, or attempts to assess and improve the computational reproducibility of published work. At the center of this process is the *reproducer* (you!), a party rarely involved in the production of the original paper. Reproductions sometimes involve the *original author* (whom we refer to as “the author”) in cases where additional guidance and materials are needed to execute the process. Reproductions should be distinguished from *replications*, where replicators re-examine a study’s hypotheses using *different data* or *different methods* (or both) (King 1995). We find that reproducibility is necessary for replicability, though both allow science to be “self-correcting.”

We recommend using this Guide in conjunction with the **Social Science Reproduction Platform** (SSRP), an open-source platform that crowdsources and catalogs attempts to assess and improve the computational reproducibility of published social science research. Though in its current version, the Guide is primarily intended for reproductions in economics, it may be used in other social science disciplines, and we welcome contributions that aim to “translate” any of its parts to other social science disciplines (learn how you can contribute here). Find definitions of fundamental concepts in reproducibility and the process of conducting reproductions in the Glossary chapter.

This Guide and the SSRP were developed as part of the Accelerating Computational Reproducibility (ACRe) project, which aims to assess, enable, and improve the computational reproducibility of published social science research. The ACRe project is led by the Berkeley Initiative for Transparency in the Social Sciences (BITSS)—an initiative of the Center for Effective Global Action (CEGA)—and Dr. Lars Vilhuber, Data Editor for the journals of the American Economic Association (AEA). This project is supported by Arnold Ventures.

View slides used for the presentation “How to Teach Reproducibility in Classwork”

Beyond binary judgments

Assessments of reproducibility can easily gravitate towards binary judgments that declare an entire paper as “(ir-)reproducible”. We suggest a more nuanced approach based on two observations that make binary judgments less relevant.

First, a paper may contain several scientific claims (or major hypotheses) that may vary in computational reproducibility. Each claim is tested using different methodologies, presenting results in one or more display items (outputs like tables and figures). Each display item will itself contain several specifications. Figure 1 illustrates this idea.

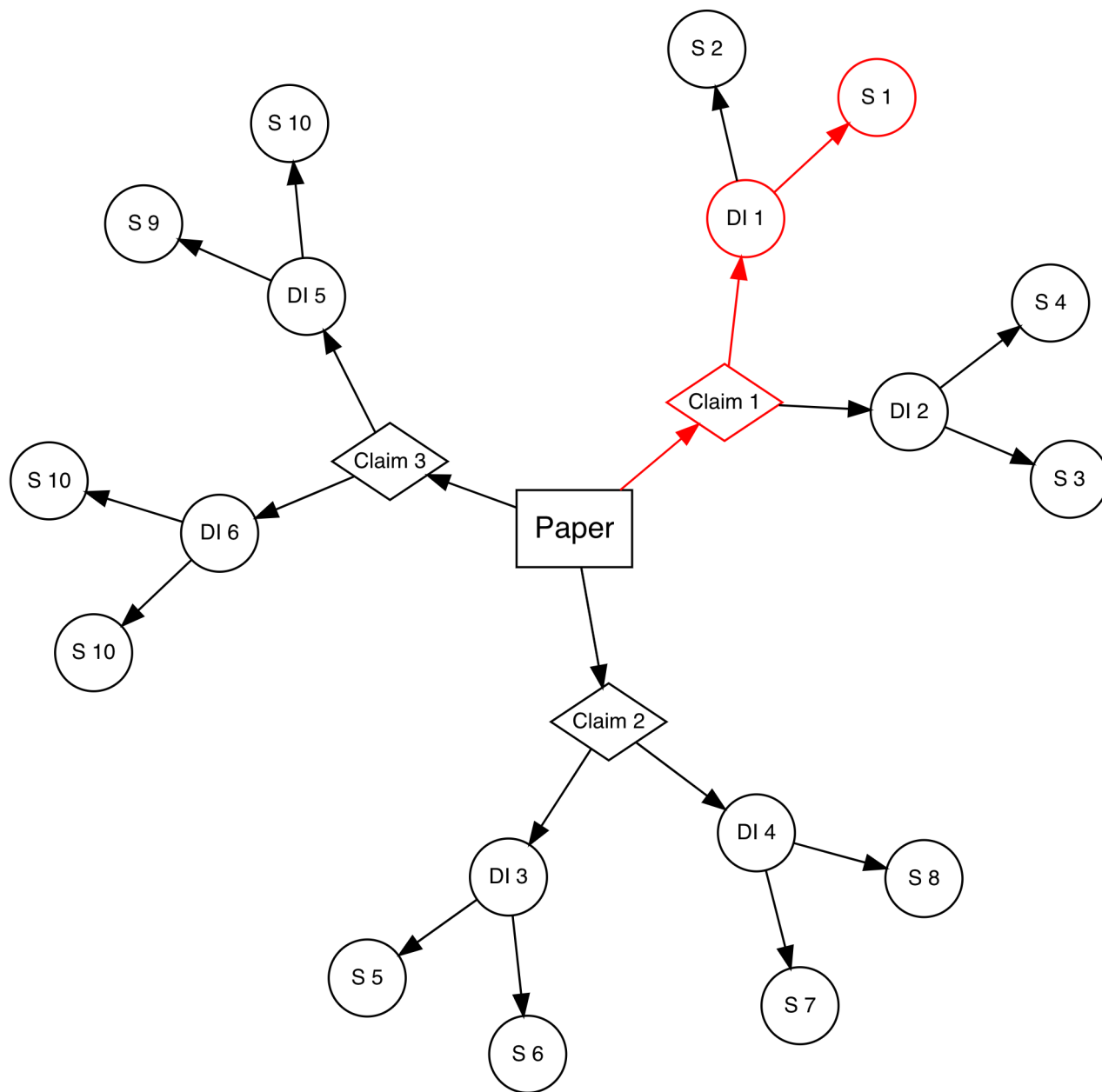


Figure 1: One paper has multiple components to reproduce.
 DI: Display Item, S: Specification

Second, for any given specification, there are several reproducibility levels, ranging from the absence of any materials to complete reproducibility starting from raw data. Moreover, even for a specific claim-specification combination, distinguishing the appropriate level can be far more constructive than simply labeling it as (ir-)reproducible.

Note that the highest level of reproducibility, which requires complete reproducibility starting from raw data, is very demanding and should not be expected of all published research — especially before 2019. Instead, this level can serve as an aspiration for social science research, as we look to improve the reproducibility of research and facilitate the transmission of knowledge throughout the scientific community.

Reproduction stages

Reproductions can be divided into five stages, corresponding to the first five chapters of this guide:

0. **Paper selection**, where you will select a *candidate* paper and try to locate its *reproduction package*. If a reproduction package is available, you will *declare* the paper and start the reproduction, or select a new candidate paper (after leaving a record on the SSRP);
1. **Scoping**, where you will define the scope of the exercise by recording the claims, display items, and specifications you will focus on in the remainder of the reproduction;
2. **Assessment**, where you will review and describe in detail the available reproduction package and assess the current level of computational reproducibility of the selected display items;
3. **Improvement**, where you will modify the content and/or the organization of the reproduction package to improve its reproducibility;
4. **Robustness**, where you will identify analytical choices to increase the number of feasible robustness checks, and/or you will justify the reasonableness of a specific robustness check.

This Guide does not include a possible fifth stage of **extension**, where you may extend the current paper by including new methodologies or data, which would bring the exercise closer to a *replication*.

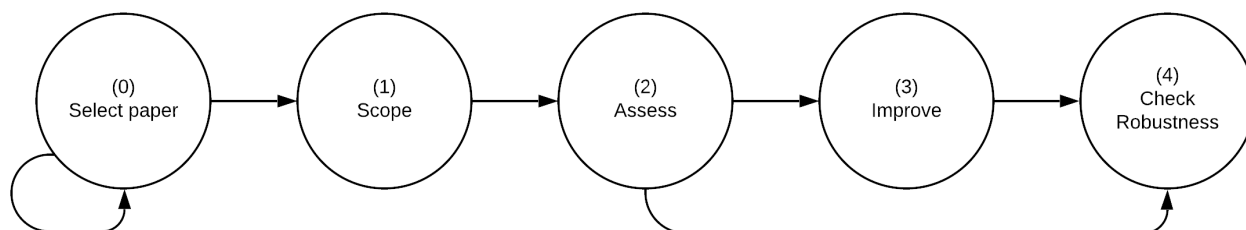


Figure 2: Four stages of a reproduction attempt

The order of the stages may not be chronologically linear. For example, you may realize that the scope of a reproduction is too ambitious and switch to a less intensive one. Later in the exercise, you can also begin testing different specifications for robustness while also assessing the paper’s reproducibility level. The only stage that should go first, and cannot be edited once finished, is the Scoping stage.

Different stages in the reproduction process correspond to different units of analysis (see Figure 3 for an overview). The Scoping stage will focus on scientific *claims* selected for reproduction. Once you specify your claims of interest, in the Assessment and Improvement stages you will focus on the display items associated with those claims. In the Robustness stage, claims are once again the unit of analysis.

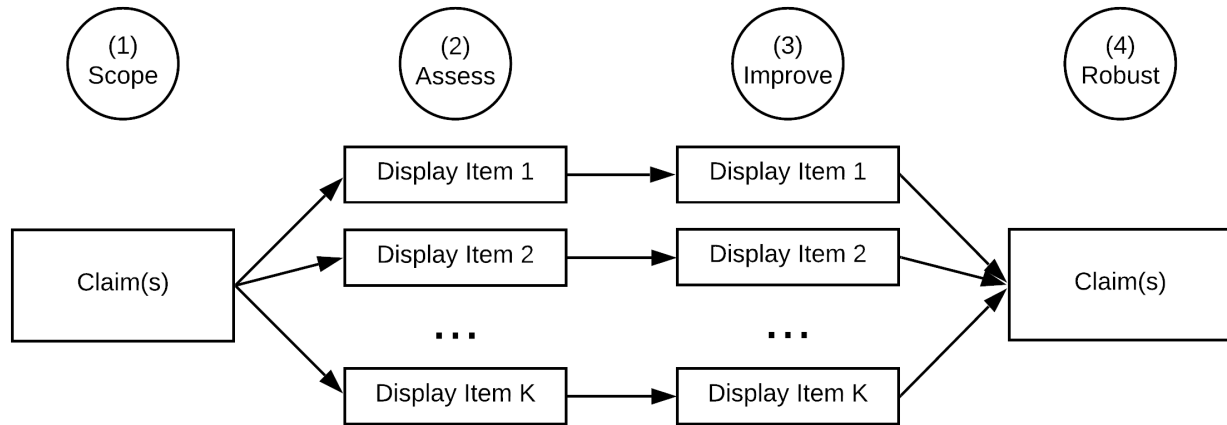


Figure 3: Relevant unit of analysis at each stage of a reproduction attempt

Reproduction strategies

In most cases, you will begin a reproduction with a thorough reading of your paper of interest. However, the sequence of the steps you take in the remainder of the reproduction may follow various reproduction strategies. The most obvious strategy would be to follow the order of the steps as outlined above. You may also first choose one of the paper's many claims and then focus on assessing and improving the reproduction package accordingly. Using an alternative strategy, you might identify potential robustness checks or extensions while reading the paper and then focus only on the results associated with that robustness check. In another strategy, you may identify a paper that uses a particular dataset in which you are interested and then only reproduce or conduct robustness checks for the results associated with that dataset.

Chapter 1

Selecting a paper

The goal of this stage is to help you select the paper that you will do your reproduction on. You might first consider multiple papers without analyzing them more closely (we refer to these as **candidate papers**) before moving forward with your **declared paper**.

The main difference between a candidate and a declared paper is the availability of a reproduction package. A **reproduction package** is the collection of materials that make it possible to reproduce a paper. This package may contain data, code, or documentation. If you are unable to independently locate the reproduction package for your paper, you can ask the paper’s author for it (find guidance on this in Chapter 7) or simply choose another candidate paper. If you still want to explore the reproducibility of a paper with no reproduction package, these guidelines provide instructions for requesting materials from authors to create a public reproduction package, or if this proves unsuccessful, for building your reproduction package from scratch.

To avoid duplicating the efforts of others who may be interested in reproducing one of your candidate papers, **we ask that you record your candidate papers in the SSRP**.

Note that in this stage, *you are not expected to review the reproduction materials in detail*, as you will dedicate most of your time to this in later stages of the exercise.

1.1 From candidate to declared paper

At this point of the exercise, you are *only verifying the availability* of (at least) one reproduction package and not assessing the quality of its content. Follow the steps below to verify that a reproduction package is available, and stop whenever you find it (this may mean that you have found your declared paper).

1. Check whether previous reproduction attempts have been recorded in the SSRP for the paper.
2. Check the journal or publisher’s website, looking for materials named “Data and Materials,” “Supplemental Materials,” “Reproduction/Replication Package/Materials,” etc.
3. Look for links in the paper (review the footnotes and appendices).
4. Review the personal websites of the paper’s author(s).
5. Contact the author(s) to request the reproduction package using this email template. In this and future interactions with authors, we encourage you to follow our guidance outlined in Chapter 7.
6. Deposit the reproduction package in a trusted repository (e.g., Dataverse, Open ICPSR, Zenodo, or the Open Science Framework) under the name **Original reproduction package for – Title of the paper**. You will be asked to provide the URL of the repository in Survey 1.

In case you need to contact the authors, make sure to *allocate sufficient time for this step* (we suggest at least three weeks before the date you plan to start the reproduction). Instructors should also plan to accordingly (e.g., if the

ACRE exercise is expected to take place in the middle of the semester, students should review candidate papers and (if applicable) contact the authors in the first few weeks of the semester).

Review the decision tree (Figure 1.1) below for a more detailed overview of this process. Remember, *if at any step of the process you decide to abandon the paper, make sure to record the candidate paper in the SSRP* before moving on to another candidate paper. Once you have obtained the reproduction package, the *candidate paper* becomes your *declared paper* and you can move forward with the exercise! Do not invest time in doing a detailed read of any paper until you are sure that it is your declared paper.

1.1.1 Candidate paper entries in the SSRP

If the SSRP contains previous reproduction attempts of the paper, you will see a report card with the following information:

Box 1: Summary Report Card for SSRP Paper Entry

Title: Sample Title

Authors: Jane Doe & John Doe

Original Reproduction Package Available: Yes (link)/No.

[If “No”] **Contacted Authors?:** Yes/No

[If “Yes(contacted)”] **Type of Response:** Categories (6).

Additional Reproduction Packages: Number (eg., 2)

Authors Available for Further Questions for SSRP Reproductions: Yes/No/Unknown

If after taking steps 1-5 above (or for some other reason) you are unable to locate the reproduction package, record your candidate paper (and if applicable, the outcome of your correspondence with the original authors) in the SSRP database following the example above.

1.2 Identifying the relevant timeline

Before you begin working on the four main stages of the reproduction exercise (Scoping, Assessment, Improvement, and Robustness), it is important to manage your own expectations and those of your instructor or advisor. Be mindful of your time limitations when defining the scope of your reproduction activity. These will depend on the type of exercise chosen by your instructor or advisor and may vary from a weeklong homework assignment, to a semester-long project (an undergraduate thesis, for example).

Table 1 shows an example distribution of time across three different reproduction formats. The Scoping and Assessment stages are expected to last roughly the same amount of time across all formats (lasting longer for the semester-long activities as less experienced researchers, such as undergraduate students, may need more time). Differences emerge in the distribution of time for the last two main stages: Improvements and Robustness. For shorter exercises, we recommend avoiding any possible improvements to the raw data (or cleaning code). This will limit how many robustness checks are possible (for example, by limiting your ability to reconstruct variables according to slightly different definitions), but it should leave plenty of time for testing different specifications at the analysis level.

1.3 Potential sources of papers to prioritize

1.3.1 Prioritize based on higher likelihood of reproduction

To curate a list of target papers in *economics*, instructors can consult recently published articles in the journals of the American Economic Association (AEA), especially articles published since 2019, which are more likely to contain reproduction packages given a change in the AEA Data and Code Availability Policy that took effect that year. In *political science*, target papers can be found in the *American Journal of Political Science*, especially those published after 2016, which are also likely to contain reproduction packages. In other social science disciplines, instructors

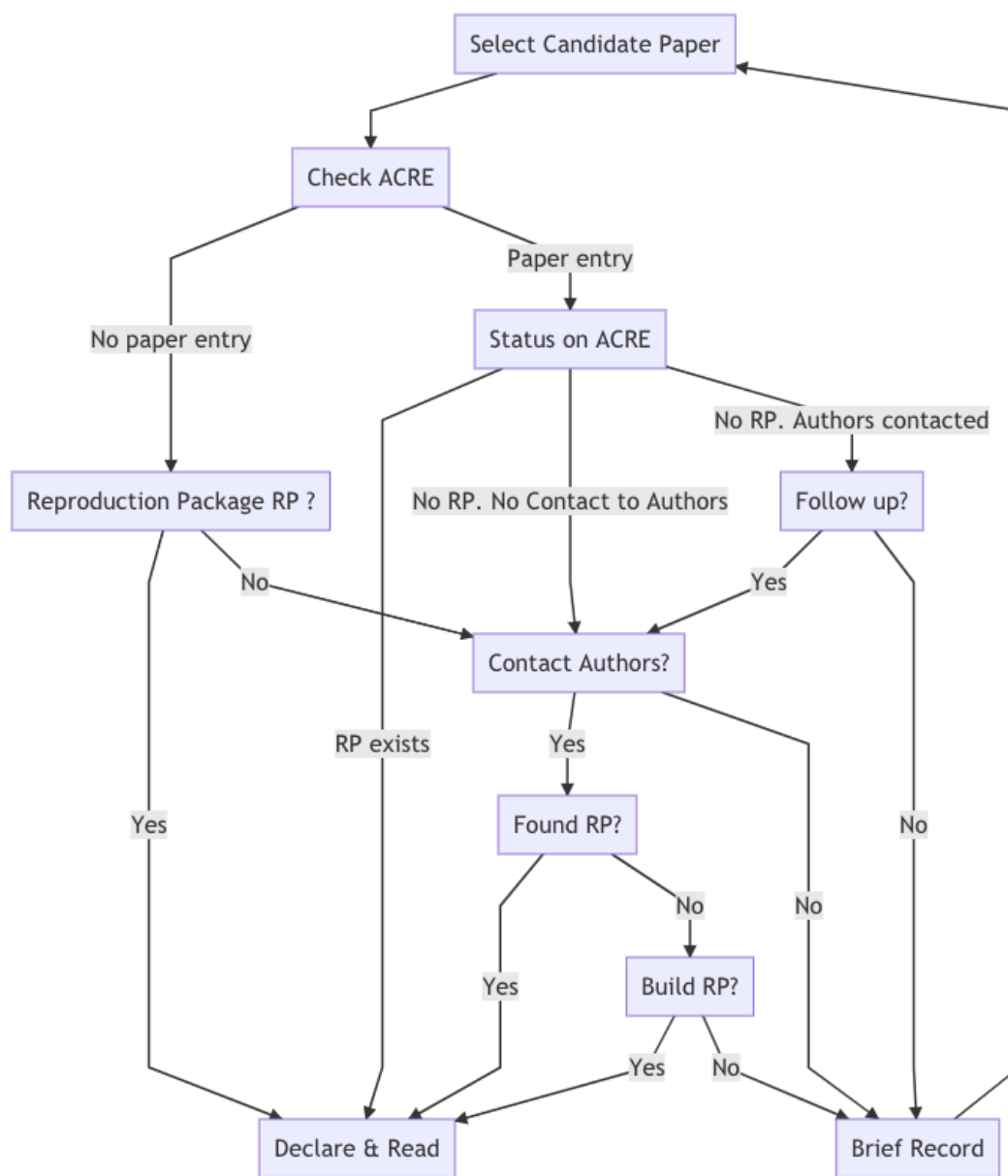


Figure 1.1: Decision tree to move from candidate to declared paper

	2 weeks (~10 days)		1 month (~20 days)		1 semester (~100 days)	
Stages of Reproduction	analysis data	raw data	analysis data	raw data	analysis data	raw data
Scoping	10% (1 day)		10% (2 day)		10% (10 days)	
Assesment	35%		25%		15%	
Improvement	25%	0%	40%		20%	30%
Robustness	25%	5%	25%		25%	

can focus on journals that score 2 or above in data, code, and materials transparency in the Transparency and Openness Promotion (TOP) Factor (see scores here), maintained by the Center for Open Science.

1.3.2 Prioritize based on proxies of paper impact within a field

Instructors and reproducers could prioritize papers based on their impact in their field. We suggest that an explicit criteria should be specified in order to define a pool of “high impact” papers. Here we demonstrate one such criteria for the field of Economics.

1. Go to Google Scholar metrics
2. Select top publications
3. Click on categories and select Business, Economics & Management. Then select in subcategories, select Economics
4. Exclude the Journal of Economics Perspectives (a review journal that does little computational analysis).
5. Focus on the 10 journals with highest h5-index. In August 4th, 2021, this yields:
 - American Economic Review
 - The Quarterly Journal of Economics
 - The Review of Financial Studies
 - Journal of Political Economy
 - The Journal of Finance
 - Econometrica
 - The Review of Economic Studies
 - Review of Economics and Statistics
 - The Economic Journal
 - Journal of Public Economics
6. For each journal click on the hyperlinked h5-index number (sort highest to lowest).
7. Select the first 10 articles in the list as sorted by citations. For example, for the *American Economic Review*, this procedure yields the following articles:
 - The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment
 - Railroads of the Raj: Estimating the Impact of Transportation Infrastructure
 - The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment
 - The Surprisingly Swift Decline of US Manufacturing Employment
 - Monetary Policy According to HANK

- The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000
 - Disruptive Change in the Taxi Business: The Case of Uber
 - Bartik Instruments: What, When, Why, and How
 - Importing Political Polarization? The Electoral Consequences of Rising Trade Exposure
 - Long-Run Impacts of Childhood Access to the Safety Net
8. Starting from the top: select a paper and read the abstract to verify that there is a computational analysis. Discard if there is not.
 9. If there is, search [socialsciencereproduction.org](https://www.socsci.uci.edu/reproduction) to verify that there is a recorded reproduction.
 10. If there is, you can create a new reproduction or move to the next paper in the list. If there is no reproduction recorded, create one.

Chapter 2

Scoping

At this stage, you will define the *scope of the reproduction* by identifying the *scientific claims* and related *display items* that you will analyze in the remainder of the reproduction. For this exercise, we follow the a comparable definition of a claim as used in the SCORE project, a related initiative aimed at predicting replicability and reproducibility of research:

“A research claim is a single major finding from a published study, as well as details of the methods and results that support this finding. A research claim [may not be] equivalent to an entire article. Sometimes the claim as described in the abstract does not exactly match the claim that is tested. In this case, you should consider the research claim to be that which is described in the [results of the paper]”.

– RepliCATS Project

In the SSRP framework, different claims in a paper may be tested using different methodologies, and their results may be presented in one or more display items, such as tables and figures (figure 0.1 illustrates this idea). Claims can have different structures, here we propose two high-level categories based on the most common structures:

- **Causal claim:** a claim is causal if it can be summarize using causal language. This language can be characterize by the following structure: “The paper estimates the effect of a variable X on outcome Y for population P , using method M ”. For example: “This paper investigates the impact of bicycle provision (X) on secondary school enrollment (Y) among young women in Bihar/India (P), using a Difference in Difference approach (M).”
- **Descriptive/predictive claim:** a claim is descriptive or predictive if it can be summarize using descriptive or predictive language. language can be characterize by the following structure: “The paper estimates the value of a variable Y (estimated or predicted) for population P under dimensions X (optional) using method M ”. For example, “Drawing on a unique Swiss data set (P) and exploiting systematic anomalies in countries’ portfolio investment positions (M), I find that around 8% of the global financial wealth of households is held in tax havens (Y).”

A *display item* is a figure or table that presents the results described in the paper. Each display item may contain several specifications, or estimates that result from variations in specific analytical choices. If the display item of interest is a figure, the specification can be recorded as an number approximated by visual inspection of a particular point in the figure.

When recording the Scoping section on the SSRP platform, count all of the claims in the paper and provide a one sentence summary for the subset of claims that you will attempt to reproduce. Structure your summaries as follows: “The paper tested the effect of X on Y for population P , using method M . The main results show an effect of magnitude E (specify units and standard errors)” or “The paper estimated the value of Y (estimated or predicted) for population P under dimensions X using method M . The main results presented an estimate of of magnitude E

(specify units and standard errors)”. Make sure to use the same units of measurement for all scientific claims that you will analyze as part of the reproduction.

Note: Once you progress past the Scoping stage on the SSRP, *you will no longer be able to edit your responses in the Scoping stage*, though you will be able to see them. This is because the content of later stages of the reproduction is dependent on the information you record at this stage of the reproduction.

2.1 Read and summarize the paper

Depending on your reproduction’s timeline, we recommend that you write a short (<1000 words) summary of the paper. Writing up such a summary will help you develop and demonstrate a wholesome understanding of the paper and its various components. In your summary, try to address the following:

- How many scientific claims can you identify in the paper?
- Would you classify the claims as causal, descriptive (e.g., estimating a population’s descriptive statistic), or something else?
- What is the population that is the focus of the paper as a whole?
- What is the population for which the estimates apply?
- What are the primary data sources used in the paper?
- What is the primary statistical or econometric method used to examine each claim?
- What is the author’s preferred specification (or yours, if the authors’ is unclear)?
- What are some possible robustness checks for the preferred specification?
- How many display items are there in the paper (tables, figures, and inline results)?

Draft the summary in a plain text editor and paste the text in the form.

2.2 Record a revised reproduction package

At the previous stage, you recorded the location of the original reproduction package made available by the paper’s authors. Given that one of the main goals of the ACRe approach is to improve reproducibility, we recommend that you build an alternative reproduction package that will improve the reproducibility of specific display items or the paper as a whole. You can start by downloading the original reproduction package (or *forking*, if on GitHub) and uploading a copy titled “Revised reproduction package for [Paper Citation, e.g. Smith et al. (2019)]” to a trusted repository. Examples of trusted repositories include Dataverse, openICPSR, Figshare, Dryad, Zenodo, Open Science Framework and others. We encourage you to also use version control software (e.g., Git) during your reproduction.

Trusted repositories have a file size limit (typically around 2gb). If you think that your reproduction package will exceed this limit please do the following:

- Separate your reproduction package in two: (1) data and (2) code and documentation.
- Post the second one in a trusted repository.
- For the data reproduction package identify all the files that are different from the original reproduction package and upload only those. For example, suppose the original repro package has `data/raw_data1.csv`, and `data/clean_data/data_set1.dta`. If you modify only the file `data_set1.dta`, then upload a revised reproduction package that has the same folder structure but only the files that differ:

`data/clean_data/data_set1.dta`. If you want to make it even more clear you could add a readme file describing the modified files.

As you work through the next stages, you can modify the reproduction package and record your improvements on the SSRP. Keeping a record of such changes will help you document your assessments or communicate with the original authors, and it will also allow future reproducers to build on top of your work. For users of version control software (e.g., Git) we recommend to track each improvement in separate commits. Then you can link the specific commits in the improvements section. This way reproducers can clearly communicate to original authors what specific lines of the code were modified (using the `diff` feature).

2.3 Record scope of the exercise

By now, you probably have a reasonably good understanding of the paper. You do not, however, need to spend any time reviewing the reproduction package in detail yet.

At this point, you should specify the parts of the paper that will be the main focus of your reproduction. Focus on specific estimates, represented by a unique combination of claim-display item-specification as represented in figure 1. Given the complexity of the SSRP approach, unless you are very familiar with the paper, we recommend starting with just one claim before moving onto second or third as part of a later reproduction.

Declare a specific estimate(s) to reproduce

Identify a scientific claim and its corresponding preferred specification and record its magnitude (specify units), standard error, and location in the paper (page, table #, and table row and column). If the authors did not explicitly choose a precise preferred estimate, you can choose one yourself. In addition to the preferred estimate, you can reproduce up to five estimates that correspond to the preferred estimate's alternative specifications. You can then enter all this information in the SSRP in a table like the one below

Note: If your display item of interest is a figure, you can record a specific point in the figure, or one that can be easily derived from it (e.g., the distance between two points). In the absence of a numerical estimate for this point, you are encouraged to record an approximate number by looking at the corresponding value in the y-axis. You should still aim at reproducing the entire figure, and report any differences in the assessment stage.

Declare possible robustness checks for main estimates (optional)

After reading the paper, you might wonder why the authors did not conduct a specific robustness test. If you think that such analysis could have been done *within the same methodology* and *using the same data* (e.g., by including or excluding a subset of the data like “high-school dropouts” or “women”), please specify a robustness test that you would like to conduct before starting the Assessment stage. Robustness checks in this stage are *optional* and can take the form of a short sentence describing at a high level what you (the reproducer) would like to explore in a later stage. In the robustness stage (after assessment and improvement) you will be able to describe in greater detail how you have modified the reproduction code.

	Preferred Specification	Alternative Spec #1
Name of display item (e.g. Table 1, Figure S3)	Table 2	Table 2
Estimate	0.169	0.088
Standard Error		0.028
Units (e.g. %, \$, log(\$), elasticity, standard deviations)	Elasticity of transp costs to	Elasticity of transp costs to
p-value		
Confidence interval	0.062,0.296	
Other Statistic		
Page	917	917
Column	2	1
Row	3	1
Inline Paragraph		
Econometric Method	Other	OLS
Specify method (if other selected)	NLS + graph theory	

Figure 2.1: SSRP table to input specifications for each claim

Chapter 3

Assessment

In the *Assessment* stage, you will describe the available reproduction materials and assign a reproducibility score to the display items associated with your selected claims. You will also review reproducibility practices for the overall paper. This stage records rich information about each reproduction to allow future reproducers to pick up easily where others have left off.

In the previous two stages, you declared a paper and identified claims and their associated estimates (found in display items) that you intend to analyze in the remainder of your reproduction. In this stage, you will get to decide whether you are interested in assessing the reproducibility of entire display items (e.g., “Table 1”) or only specific estimates found in display items (e.g., “rows 3 and 4 of Table 1”). You can also include additional display items of interest.

The Assessment stage aims to analyze the reproduction package’s *current* reproducibility—before you suggest any improvements. By the end of this section, you will have created a highly detailed description of the reproduction package’s current reproducibility that you can use to implement improvements, potentially with the paper’s original authors’ help.

On the SSRP, you will first provide a detailed description of available inputs in the reproduction package. You will then connect the display items you’ve chosen to reproduce with their corresponding inputs. With these elements in place, you can assign a score of each display item’s reproducibility and record various paper-level dimensions of reproducibility.

Tip: We recommend that you first focus on just one display item (e.g., “Table 1”). After completing the assessment for this display item, you will have a much easier time assessing others.

3.1 Describe the inputs

This section explains how to record the *input* materials found (or referenced) in the reproduction package. At this point, it may be challenging to precisely identify the materials that correspond to your selected claims and display items, so we recommend listing *all* files in the reproduction packages (*tip:* using the command line, go to the directory of the reproduction package, and type `file */*` on a Mac or `dir /s /b /o:gn` on Windows to obtain a printout of all files within a folder). However, if the reproduction package is too extensive to record in its entirety, you can focus only on the materials required to reproduce a specific display item.

In section “Describe input” on the SSRP, you can record data sources and connect them with their raw data files (if available). You can then locate and provide a brief description of the analytic data files and then record the inputs and outputs of each script file.

Tip: For large volume of data sources, we recommend downloading the empty table as a CSV file, recording the information in the downloaded file, and uploading the final version of the document to populate the table.

The following concepts may be helpful as you work in the Assessment stage (also see the ACRé glossary here):

- **Raw data:** Unmodified data files obtained by the authors from the sources cited in the paper. Data from which personally identifiable information (PII) has been removed are *still considered raw*. All other modifications to raw data make it *processed*.
- **Analysis/Analytic data:** Data used as the final input in a workflow to produce a statistic displayed in the paper (including appendices).
- **Cleaning code:** A script associated primarily with data cleaning. Most of its content is dedicated to actions like deleting variables or observations, merging data sets, removing outliers, or reshaping the data structure (from long to wide, or vice versa).
- **Analysis code:** A script associated primarily with analysis. Most of its content is dedicated to generate some type of estimate, numerical or visual, to be presented in the paper (including the appendix). Examples of computations that lead to such estimates are: running regressions, running hypothesis tests, computing standard errors, and imputing missing values, generating a series and plotting it.

3.1.1 Describe the data sources and raw data

In the paper you chose, find references to all *data sources* used in the analysis. A data source is usually described in narrative form. For example, if in the body of the paper, or the appendix, you see text like “(...) for earnings in 2018 we use the Current Population Survey”, the data source should be recorded as “2018 Current Population Survey”. If the first reference to this data source is found on page 1 of the appendix, you should record its location as “A1”. Do this for all data sources mentioned in the paper. Each row represents a unique data source.

Data sources also vary by unit of analysis, with some sources matching the same unit of analysis used in the paper (as in previous examples). In contrast, others may be less clear, e.g., “our information on regional minimum wages comes from the Bureau of Labor Statistics.” You should record such data source as “regional minimum wages from the Bureau of Labor Statistics.”

Next, look at the reproduction package and map the *data sources* mentioned in the paper to the *data files* in the reproduction package. In the **Location** column, record their folder locations relative to the main reproduction folder¹. In addition to looking at the existing data files, we recommend that you also review the first lines of all code files (especially cleaning code), looking for lines that call the datasets. Inspecting these scripts may help you understand how different data sources are used, and possibly identify any missing files from the reproduction package. Whenever a data source contains multiple files, enter them in the same cell, separated by semicolon (;).

If you cannot find the files names corresponding to a specific data source, type “Not available” in the **Data Files** field. If you can identify the file name. Check the **Provided** column if the data source was included in the original reproduction package. Check the **Cited** column if the data source was explicitly cited in the paper. Record your work using the following structure:

Table 3.1: Raw data information

Data.Source	Page	Data.Files	Location	Provided	Cited
“Current Population Survey 2018”	A1	cepr_march_2018.dta	data/	TRUE	FALSE
“Provincial Administration Reports”	A4	coast_simplepoint2.csv; rivers_simplepoint2.csv; ...	Data/maps/	TRUE	FALSE
“2017 SAT scores”	4	Not available	data/to_clean/	FALSE	TRUE
...

Note on data citations: Assessment of whether or not a data source has been cited should take into account the general guidelines for what is considered a complete data citation. In-text citations, mention of the source location (e.g. url to source location), description of the source or the data alone are examples of incomplete data citations. Please refer to the Guidance on Data Citations by Social Science Data Editors for a comprehensive overview and more reference material on data citation standards for the social sciences.

¹a relative location takes the form of `folder_in_rep_materials/sub_folder/file.txt`, in contrast to an absolute location that takes the form of `/username/documents/projects/repros/folder_in_rep_materials/sub_folder/file.txt`

3.1.2 Describe the analytic data

First, identify all analytic data files in the reproduction package and record their names in the **Analytic Data** column in Table 1.2. You will recognize analytic data files based on the documentation, their location folder, or if a code file either uses them to compute a statistic that is displayed in the paper (or appendix).

Second, record each analytic data file’s location relative to the main folder of the reproduction package in the **Location** column.

Finally, provide a one-line description of each file in the **Description** column (e.g., `all_waves.csv` can be “data for region-level analysis”). This will become easier as you progress through the reproduction, and you can always return to this table later on and add or modify your descriptions.

The resulting table may look like this:

Table 3.2: Analysis data information

Analytic.Data	Location	Description
<code>final_data.csv</code>	<code>analysis/fig1/</code>	data for figure1
<code>all_waves.csv</code>	<code>final_data/v1_april/</code>	data for region-level analysis
...

3.1.3 Describe the code scripts

First, identify all code files in the reproduction package and record their names in the **File Name** column and record their locations relative to the main folder in the **Location** column.

Then, review the beginning and end of each code file to identify the inputs required to successfully run the file and the outputs it produces. Inputs are data sets or other code scripts that are typically found at the beginning of the script (e.g., `load`, `read`, `source`, `run`, `do`). Outputs are other data sets, or plain text files typically found at the end of a script (e.g., `save`, `write`, `export`). Record those in the *Inputs* and *Outputs* columns.

Finally, provide a brief description of the code’s function in the *Description* column and classify its function as analysis or cleaning and/or construction in the *Primary Type* column.

For each code file, record all outputs together and separate each item with (;). Record all of this information in the SSRP using the following structure:

Table 3.3: Code files information

file_name	location	inputs	outputs	description	primary_type
<code>output_table1.do</code>	<code>code/analysis/</code>	<code>analysis_data01.csv</code>	<code>output1_part1.txt</code>	produces first part of table 1 (unformatted)	analysis
<code>data_cleaning02.R</code>	<code>code/cleaning</code>	<code>admin_01raw.csv</code>	<code>analysis_data02.csv</code>	removes outliers and missing vals from raw admin data	cleaning
...

Reproducers are encouraged to build on top of previous reproductions by exporting input descriptions of previous reproducers (by searching into previous reproductions and clicking “download table as csv” in the assessment section) and uploading them into their own reproduction (by clicking “upload csv”).

3.1.3.1 Simplify several files with a recursive structure

If there are many files for a given data source, analysis data or code scripts, that have a recursive structure (e.g., `wages0001.csv`, ... , `wages0734.csv`), record only one generic entry that represents all these of files (e.g., `wages0XYZ.csv`).

3.2 Connect display items to all its inputs using the Diagram Builder

Using the information above, the SSRP will build a diagram that can help you visually trace your display items to the code and data files that produce them.

3.2.1 Complete workflow information

If you were able to identify all of the relevant components in the previous section, and if the code scripts are have the display items as final outputs, the Diagram Builder will produce a tree diagram that looks similar to the one below.

```

table1.tex
  [code] analysis.R
    analysis_data.dta
      [code] final_merge.do
        cleaned_1_2.dta
          | [code] clean_merged_1_2.do
          | merged_1_2.dta
          | [code] merge_1_2.do
          | cleaned_1.dta
          | | [code] clean_raw_1.py
          | | raw_1.dta
          | cleaned_2.dta
          | [code] clean_raw_2.py
          | raw_2.dta
        cleaned_3_4.dta
          [code] clean_merged_3_4.do
            merged_3_4.dta
              [code] merge_3_4.do
                cleaned_3.dta
                | [code] clean_raw_3.py
                | raw_3.dta
                cleaned_4.dta
                [code] clean_raw_4.py
                raw_4.dta

```

This diagram, built with the information you provided, contributes to understanding the necessary components required to reproduce a specific display item. It also summarizes key information to allow for more constructive exchanges with original authors or other reproducers. For example, when contacting the authors for guidance, you can use the diagram to ask for specific files. Formulating your request this way makes it easier for authors to respond and demonstrates that you understand the reproduction package. You can also add this diagram to the readme of your revised reproduction package.

3.2.2 Incomplete workflow information

In some cases, some of the workflow components may not be easily identifiable in the reproduction package (or might be missing). Here the Diagram Builder will return a fragmented reproduction tree diagram. See below how you can work around such cases, but even if the reproduction tree is fragmented, you can still go on to the next step of the reproduction.

For example, here's a simple complete reproduction tree:

```

table1.tex
  analysis.R

```

```

analysis_data.dta
  final_merge.do
    cleaned_1_2.dta
      clean_merged_1_2.do
        merged_1_2.dta

```

For this case, if the file `final_merge.do` is missing, the Diagram Builder will produce the following reproduction tree:

```

table1.tex
  analysis.R
    analysis_data.dta

cleaned_1_2.dta
  clean_merged_1_2.do
    merged_1_2.dta

```

You can still manually combine this partial information with your knowledge from the paper and own (subjective) judgment to produce a “candidate” reproduction tree. This may look like the following:

```

table1.tex
  analysis.R
    cleaned_1_2.dta
      MISSING_CODE_FILE_1
        cleaned_1_2.dta
          clean_merged_1_2.do
            merged_1_2.dta

```

If you notice that the reproduction tree is fragmented where it shouldn’t be, you may need to amend Table 1.3 above with placeholders. In the example above, that can look like this:

Table 3.4: Adding rows to code spreadsheet

file_name	location	inputs	outputs	description	primary_type
...
MISSING_ CODE FILE 1	unknown	cleaned_1_2.dta	cleaned_1_2.dta	missing code	unknown

As in the cases with complete workflows, these diagrams (fragmented or reconstructed trees) provide important information for assessing and improving the reproducibility of specific display items. For more examples of diagrams connecting final outputs to initial raw data, see here.

3.2.2.1 Common issues that might occur when creating a reproduction tree

- **Script output is not a display item.** Sometimes the reproduction package will not produce display items as its final outputs. In this situation, the final code script will generate some type of output (e.g., `results1.log`, `results2.csv`) that will require manual copying and pasting to reproduce the desired display item (e.g. Table 1). In this case, we recommend adding one auxiliary line to the table linking the final output to the desired display item (e.g., File Name: `aux1`, Inputs: `results1.log`; `results2.csv`, Outputs: Table 1).
- **Loops across files.** If you notice that some files iterate between each other (e.g., `file1.do` calls `data1.csv` to generate `data2.csv`, and `file2.R` calls `data2.csv` to generate `data1.csv`), look within the files to identify the one that contains some stopping criteria (e.g., stop when SSR is minimized). Then designate only one output as the final and record that in reproduction diagram in the next section.

- **Large number of recurrent files.** Some reproduction packages might have a large number of files that follow a recurrent structure (e.g., `data_location1_year1`, `data_location2_year1`, `data_location1_year2`, ...). To make a more tractable tree we recommend that you record the names of these files only once and use a consistent place holder (e.g., `data_locationX_yearY`).
- **Possible bugs due to missing ;.** If the reproduction tree shows several files under one input, make sure that each file is separated by a semi-colon (;), that way the tree will separate each file into a separate input.

3.3 Assign a reproducibility score.

Once you have identified all possible inputs and have a clear understanding of the connection between the display items and their inputs, you can assign reproducibility scores to individual display items.

The following concepts may be helpful in this section:

- **Computationally Reproducible from Analytic data (CRA)** — The output can be reproduced with minimal effort starting from the *analytic* datasets.
- **Computationally Reproducible from Raw data (CRR)** — The output can be reproduced with minimal effort from the *raw* datasets.
- ****Standard of *minimal effort**** — One hour or less is required to run the code, not including computing time.

3.3.1 Levels of Computational Reproducibility for a Specific Display Item

Each level of computational reproducibility is defined by the availability of data and materials, and whether the available materials faithfully reproduce the display item of interest. The description of each level also includes possible improvements that can help advance the display item's reproducibility. You will learn in more detail about the possible improvements at the *Improvement* stage.

Note that the assessment is made *at the level of individual display items*—a paper can be highly reproducible for its main results, but its other display items may not be as reproducible. The assessment includes a 10-point scale, where 1 represents that, under the current circumstances, reproducers cannot access any reproduction package. At level 10, the reproducer can access all of the necessary materials to faithfully reproduce the target display item from the raw data.

- **Level 1 (L1):** No data or code are available. Possible improvements include adding: raw data, analysis data, cleaning code, and analysis code.

You may have detected papers without any reproduction materials at the *Paper Selection* stage, where you should have recorded them as abandoned candidate papers.

- **Level 2 (L2):** Code scripts are available (partial or complete), but no data are available. Possible improvements include adding: raw data and analysis data.
- **Level 3 (L3):** Analytic data and code are partially available, but raw data and cleaning code are missing. Possible improvements include: completing analysis data and/or code, adding raw data, and adding analysis code.
- **Level 4 (L4):** All analytic data sets and analysis code are available, but the code fails to run or produces results inconsistent with the paper (not CRA). Possible improvements include: debugging the analysis code or obtaining raw data.
- **Level 5 (L5):** Analytic data sets and analysis code are available and they produce the same results as presented in the paper (CRA). The reproducibility package may be improved by obtaining the original raw data.

Note: This is the highest level that most published research papers can attain currently. Computational reproducibility *from raw data* is required for papers that are reproducible at Level 6 and above.

- **Level 6 (L6):** Cleaning code scripts are available (partial or complete), but raw data is missing. Possible improvements include: adding raw data.
- **Level 7 (L7):** Cleaning code is available and complete, and raw data is partially available. Possible improvements: adding raw data.
- **Level 8 (L8):** All the materials (raw data, analytic data, cleaning code, and analysis code) are available. However, the cleaning code fails to run or produces different results from those presented in the paper (not CRR) or the analysis code fails to run or produces results inconsistent with the paper (not CRA). Possible improvements: debugging the cleaning or analysis code.
- **Level 9 (L9):** All the materials (raw data, analytic data, cleaning code, and analysis code) are available. The analysis code produces the same output as presented in the paper (CRA). However, the cleaning code fails to run or produces different results from those presented in the paper (not CRR). Possible improvements: debugging the cleaning code.
- **Level 10 (L10):** All necessary materials are available and produce consistent results with those presented in the paper. The reproduction involves minimal effort and can be conducted starting from the analytic data (CRA) and the raw data (CRR). *Note that Level 10 is aspirational and may be unattainable for most research published today.*

The following table summarizes the different levels of computational reproducibility (for any given display item). For each level, there are reproducibility components and practices that are present () or can be implemented to advance the level of reproducibility (-).

Table 3.5: Levels of Computational Reproducibility
(P denotes “partial”, C denotes “complete”)

	Availability of materials, and reproducibility									
	Analysis Code		Analysis Data		CRA	Cleaning Code		Raw Data		CRR
	P	C	P	C		P	C	P	C	
L1: No materials	–	–	–	–	–	–	–	–	–	–
L2: Only code	✓	✓	–	–	–	–	–	–	–	–
L3: Partial analysis data & code	✓	✓	✓	–	–	–	–	–	–	–
L4: All analysis data & code	✓	✓	✓	✓	–	–	–	–	–	–
L5: Reproducible from analysis	✓	✓	✓	✓	✓	–	–	–	–	–
L6: All cleaning code	✓	✓	✓	✓	–	✓	✓	–	–	–
L7: Some raw data	✓	✓	✓	✓	–	✓	✓	✓	–	–
L8: All raw data	✓	✓	✓	✓	–	✓	✓	✓	✓	–
L9: All raw data + CRA	✓	✓	✓	✓	✓	✓	✓	✓	✓	–
L10: Reproducible from raw data	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
^a Computationally Reproducible from Analytic data (CRA): The output can be reproduced with minimal effort starting from the <i>analytic</i> datasets.										
^b Computationally Reproducible from Raw data (CRR): The output can be reproduced with minimal effort from the <i>raw</i> datasets.										

Figure 3.1: Table 3.5.

You may disagree with some of the levels outlined above, particularly wherever subjective judgment may be required. If so, you are welcome to interpret the levels as unordered categories (independent from their sequence) and suggest improvements using the “Edit” button above (top left corner if you are reading this document in your browser).

Note: The levels suggested here are to be used in the Assessment stage and, optionally, for the Improvement stage. *These levels are not meant to be used in the Robustness stage.*

3.3.2 Adjusting Levels To Account for Confidential/Proprietary Data

A large fraction of published research in the social sciences uses confidential or proprietary data (e.g., government data from tax records or service provision), generally referred to as *administrative data*. Since administrative data are rarely readily publicly available, some reproducibility levels presented above only apply once modified. The underlying theme of these modifications is that when data cannot be provided, you can assign a reproducibility score based on the level of detail in the instructions for accessing the data, typically provided in data availability statements (DAS). Use this checklist (.pdf, .md) to guide the assessment, and potential improvement, of a DAS.

When the reproducer cannot directly assess the reproducibility based on publicly available materials, the reproduction package should include certification that a third party (not involved in the paper's production) faithfully reproduced the results.

- **Levels 1 and 2** can be applied as described above.
- **Adjusted Level 3 (L3*)**: All analysis code is provided, but only partial instructions to access the *analysis data* are available. This means that the original authors have provided some, but not all, of the following information:
 - a. *Contact information*, including name of the organization(s) that provides access to at least one individual's data and contact information.
 - b. *Terms of use*, including licenses and eligibility criteria for accessing the data, if any.
 - c. *Information on data files (meta-data)*, including the name(s) and number of files, file size(s), relevant file version(s), and number of variables and observations in each file. Though not required, other relevant information may be included, including a dataset dictionary, summary statistics, and synthetic data (fake data with the same statistical properties as the original data).
 - d. *Estimated costs for access*, including monetary costs such as fees and licenses required to access the data, and non-monetary costs such as wait times and specific geographical locations from where researchers need to access it.
- **Adjusted Level 4 (L4*)**: All analysis code is provided, and complete and detailed instructions on how to access the *analysis data* are available.
- **Adjusted Level 5 (L5*)**: All requirements for Level 4* are met, and the authors provide a certification that a third party was able to reproduce the display item (or the paper as a whole) from the analysis data (CRA). Such certification may include a signed letter by a disinterested reproducer or a certificate from a certification agency for data and code (e.g., see CASCAD).
- **Levels 6** can be applied as described above.
- **Adjusted Level 7 (7*)**: All requirements for Level 6* are met, but instructions for accessing the *raw data* are incomplete. Use the instructions described in Level 3 above to assess the instructions' completeness.
- **Adjusted Level 8 (L8*)**: All requirements for Level 7* are met, and instructions for accessing the *raw data* are complete.
- **Adjusted Level 9 (L9*)**: All requirements for Level 8* are met, and a certification is provided that the display item can be reproduced from the analysis data (CRA).
- **Adjusted Level 10 (L10*)**: All requirements for Level 9* are met, and a certification is provided that the display item can be reproduced from the raw data (CRR).

3.3.3 Reproducibility dimensions at the paper-level

There are many tools and practices that facilitate the computational reproducibility of the paper as a whole. You can learn more about implementing such reproducibility tools and practices in the *Improvements* stage, but at this stage, you should only verify whether the current reproduction materials make use of any such tools and practices.

Note: The Assessment stage is the minimum requirement to submit your reproduction. To gain a better understanding of the paper and to help improve the reproducibility of social science research, however, we encourage you to also complete the *Improvements* and *Robustness* stages.

Table 3.6: Levels of Computational Reproducibility with Proprietary/Confidential Data
(P denotes "partial", C denotes "complete")

	Availability of materials, and reproducibility									
	Analysis Code		Instr. Analysis Data		CRA	Cleaning Code		Instr. Raw Data		CRR
	P	C	P	C		P	C	P	C	
L1: No materials	–	–	–	–	–	–	–	–	–	–
L2: Only code	✓	✓	–	–	–	–	–	–	–	–
L3: Partial analysis data & code	✓	✓	✓	–	–	–	–	–	–	–
L4*: All analysis data & code	✓	✓	✓	✓	–	–	–	–	–	–
L5*: Proof of third party CRA	✓	✓	✓	✓	✓	–	–	–	–	–
L6: All cleaning code	✓	✓	✓	✓	–	✓	✓	–	–	–
L7*: Some instr. for raw data	✓	✓	✓	✓	–	✓	✓	✓	–	–
L8*: All instr. for raw data	✓	✓	✓	✓	–	✓	✓	✓	✓	–
L9: L8 + Proof of third party CRA	✓	✓	✓	✓	✓	✓	✓	✓	✓	–
L10*: Proof of third party CRR	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

^a **Computationally Reproducible from Analytic data (CRA):** The output can be reproduced with minimal effort starting from the *analytic* datasets.

^b **Computationally Reproducible from Raw data (CRR):** The output can be reproduced with minimal effort from the *raw* datasets.

Figure 3.2: Table 3.6.

Chapter 4

Improvements

As you assess a paper, you can start proposing ways to improve its reproducibility. These improvements can be at the paper level or specific to a display item. The Social Science Reproduction Platform (SSRP) also allows you to record improvements that you’ve already implemented or that you suggest for future reproducers (including yourself) to implement. Considering improvements is an opportunity to gain a deeper understanding of a paper’s methods, findings, and overall contributions. Each contribution can also be assessed and used by the wider SSRP community, including other students and researchers using the SSRP.

As with the *Assessment* stage, we recommend that you first focus on one specific display item (e.g., “Table 1”). After making improvements to this first item, you will have a much easier time translating those improvements to other items.

4.1 Display item improvements

As part of your assessment of specific display items, you will identify potential issues with the original reproduction package. In addition to identifying these gaps, you are encouraged to implement specific improvements. In this section we suggest steps on how to add missing materials (data or code), or debug analysis or cleaning code. Record these improvements in the “Display item improvements” section.

4.1.1 Adding raw data: missing files or metadata

Reproduction packages often do not include all original raw datasets. To obtain any missing raw data or information about them, follow these steps:

1. Identify the missing file. During the Assessment stage, you identified all data sources from the paper’s body and appendices (step 1.1). However, some data sources (as collected by the original investigators) might be missing one or more files. You can sometimes find the specific name of those files by looking at the beginning of the cleaning code scripts.
2. Verify whether this file (or files) can be easily obtained from the web.
 - 2.1 - If yes: obtain the missing files and add them to your revised reproduction package. Make sure to obtain permission from the owners of this data source to publicly share this data. See chapter 7 for more guidance.
 - 2.2 - If no: proceed to step 3.
3. Eventually you will be able to use the SSRP to verify whether previous reproducers have contacted the authors regarding this paper and the specific missing files. For now, skip to the next step.

4. Contact the original authors and politely request the original materials. Be mindful of their time, and remember that the paper you are trying to reproduce was possibly published at a time when standards for computational reproducibility were different. See chapter 7 for sample language on how to approach the authors for this specific scenario.
5. If the datasets are not available due to legal or ethical restrictions, you can still improve the reproduction package by providing detailed instructions on how to access these data. For future researchers to follow, including contact information and possible costs of obtaining the raw data (e.g., access fees, how much time it might take between requesting and receiving access, etc.). Use this checklist (.pdf, .md) as a template to fill in.

4.1.2 Adding missing analytic data files

Analytic data might be missing for two reasons: (1) raw data exists, but the procedures to transform it into analytic data are not fully reproducible, or (2) some or all raw data is missing, and some or all analytic data is not included in the original reproduction package. To obtain any missing analytic data, follow these steps:

1. Identify the specific name of the missing data set. Typically this information can be found in some of the analysis code that calls the data to perform an analysis (e.g., `analysis_data_03.csv`).
2. Verify that the data cannot be obtained by running the data cleaning code over the raw data.
3. Eventually you will be able to use the SSRP to verify if previous attempts have been made to contact the authors about this data. For now, skip to the next step.
4. Contact the authors and request the specific data set.

4.1.3 Adding missing analysis code

Analysis code can be added when analytic data files are available, but some or all methodological steps are missing from the code. In this case, follow these steps:

1. Identify the specific line or paragraph in the paper that describes the analytic step that is missing from the code (e.g., “We impute missing values to...” or “We estimate this regression using a bandwidth of...”).
2. Identify the code file and the approximate line in the script where the analysis can be carried out. If you cannot find the relevant code file, identify its location relative to the main folder using the steps in the reproduction diagram.
3. Eventually you will be able to use the SSRP to verify if previous attempts have been made to contact the authors about this issue. For now, skip to the next step.
4. Contact the authors and request the specific code files.
5. If step #4 does not work, we encourage you to attempt to recreate the analysis using your own interpretation of the paper, and making explicit your assumptions when filling in any gaps.

4.1.4 Adding missing data cleaning code

Data cleaning (processing) code might be added when steps are missing in the creation or re-coding of variables, merging, subsetting of the data sets, or other steps related to data cleaning and processing. You should follow the same steps you used when adding missing analysis code (steps 1-5 above).

4.1.5 Debugging analysis code

Whenever code is available in the reproduction package, you should be able to debug those scripts. There are at least five types of debugging that can improve the reproduction package:

- *Code cleaning*: Simplify the instructions (e.g., by wrapping repetitive steps in a function or a loop) or remove redundant code (i.e., old code that was commented out) while keeping the original output intact.
- *Performance improvement*: Replace the original instructions with new ones that perform the same tasks but take less time (e.g., choose one numerical optimization algorithm over another while still obtaining the same results).
- *Adding unit tests*: Add `if/then` statements after a code chunk or section (more or less every 100 lines of code) where you test that variables or statistics are computing as expected. If the value has changed, create a warning message that mentions what object has changed.
- *Environment set up*: Modify the code to include correct paths to files, specific versions of software, and instructions to install missing packages or libraries.
- *Correcting errors*: A coding error will occur when a section of the code in the reproduction package executes a procedure that is in direct contradiction with the intended procedure expressed in the documentation (i.e., paper or code comments). For example, an error will occur if the paper specifies that the analysis is performed on a population of males, but the code restricts the analysis to females only.

4.1.6 Debugging cleaning code

Follow the same steps that you did to debug the analysis code (above), but report them separately.

4.1.7 Adding information on how to access confidential/proprietary data

If the original authors are unable to share the raw or analytical data due to legal or ethical reasons, the reproduction package can still be improved by including information on how to access such data. The AEA Data and Code Availability Policy requires authors to include data availability statements (DAS) in their README files. Data availability statements include information on “how, where, and under what conditions an independent researcher can access the original source data, as well as author-generated derivative data, and must be explicit and accurate about any restrictions, requirements, payments, and processing delays.”

Use this form (.pdf, .md) to improve the completeness of the paper’s current DAS (if any), and upload it to your revised reproduction package.

4.2 Paper-level improvements

There are several measures you can take to improve a paper’s overall reproducibility. These additional improvements can be applied across all reproducibility levels (including level 10). Record these improvements in the “Paper-level improvements” section of the SSRP.

File documentation and organization:

1. Set up the reproduction package using version control software, such as Git.
2. Improve documentation by adding comments to the code.
3. Re-organize the reproduction package into a set of folders and sub-folders that follow standardized best practices, and add a master script that executes all the code in order, with no further modifications. See AEA’s reproduction template.

Computation:

4. Integrate the documentation with the code by adapting the paper into a literate programming environment (e.g., using Jupyter notebooks, RMarkdown, or a Stata Dynamic Doc).
5. If the code was written using proprietary statistical software (e.g., Stata or Matlab), re-write some parts of it using open-source statistical software (e.g., R, Python, or Julia).
6. Set up a computing capsule that executes the entire reproduction in a web browser without needing to install any software. For examples, see Binder and Code Ocean.

Please suggest other paper-level improvements by editing this guide (use the “edit” button above) or contacting acre+feedback@berkeley.edu.

4.3 Documenting the improvements using version control

When reporting your improvements in the SSRP, we suggest using version control software (git) to track the differences between the original reproduction package and your proposed improvements. One possible approach could be the following:

1. Create an empty repository for your revised reproduction package.
2. Deposit the original reproduction package in this repository, then commit this changes using the name “depositing original reproduction package”.
3. In order to clearly show where are your improvements relative to the original reproduction package you then can take one of the following strategies:
 - 3a. Spaced commits. Wait until you are confident to have produce a concrete improvement and then commit. Or
 - 3b. Commit as often as you want, but provide the identifiers (tags) of two commits: one for to mark the reproduction package *before* you initiate a specific change (e.g., adding missing analytic data), and a second commit with the reproduction package that contains the final version of this specific improvement. With this two identifiers, readers of your reproduction will be able to easily compare (make `diffs` in git) to see exactly what was added and/or deleted.
4. Refer to the this specific commits (their tags) when describing a specific improvement in the SSRP.

Chapter 5

Checking for Robustness

Once you have assessed and potentially improved the computational reproducibility of the display items for a claim within a paper, you can determine these results' robustness by modifying some analytic choices and reporting their subsequent effects on the estimates of interest, i.e., conducting *robustness checks*. The universe of robustness checks can be very large (potentially infinite!) and pertain to data analysis and data cleaning. The SSRP distinguishes between *reasonable* and *feasible* robustness checks.

Reasonable robustness checks (Simonsohn et. al., 2018) are (i) sensible tests of the research question, (ii) expected to be statistically valid, and (iii) not redundant with other specifications in the set. The set of **feasible robustness checks** is defined by all the specifications that can be computationally reproduced. We assume that the specifications already published in the paper are part of the set of reasonable specifications.

The size of the set of feasible robustness checks, and the likelihood that it contains reasonable specifications, will depend on the current level of reproducibility of the results that support a claim. The idea is illustrated in Figure 5.1. At Levels 1-2, it is impossible to perform additional robustness checks because there are no data to work with. It may be possible to perform additional robustness checks for claims supported by display items reproducible at Levels 3-4, but not using the estimates declared in *Scoping* (because the display items are not computationally reproducible from analysis data). It is possible to conduct additional robustness checks to validate the core conclusions of a claim based on a display item reproducible at Level 5. Finally, claims associated with display items reproducible at Level 6 or higher allow for robustness checks that involve data cleaning, like variable definitions and other types of analytical choices, available only from the raw data.

The number of feasible robustness checks grows exponentially with improved reproducibility. For example, when checking the robustness of a new variable definition, you could test alternative variable definitions *and* changes in the estimation method using such variable.

Robustness is assessed at the claim level (see the diagram with a typical paper's components 1). There may be several specifications presented in the paper for a given claim, one of which the authors (or you, if the authors did not indicate) have chosen as the main or preferred specification. Identify which display item contains this specification and refer to the reproduction tree to identify the code files in which you can modify a computational choice. Using the example tree discussed in the *Assessment* stage, we can obtain the following (we removed the data files for simplicity). This simplified tree provides a list of potential files in which you can test different specifications:

```
table1.tex (contains preferred specification of a given claim)
|___[code] analysis.R
|___[code] final_merge.do
|___[code] clean_merged_1_2.do
|       |___[code] merge_1_2.do
|       |       |___[code] clean_raw_1.py
|       |       |___[code] clean_raw_2.py
|___[code] clean_merged_3_4.do
|       |___[code] merge_3_4.do
|       |       |___[code] clean_raw_3.py
|       |       |___[code] clean_raw_4.py
```

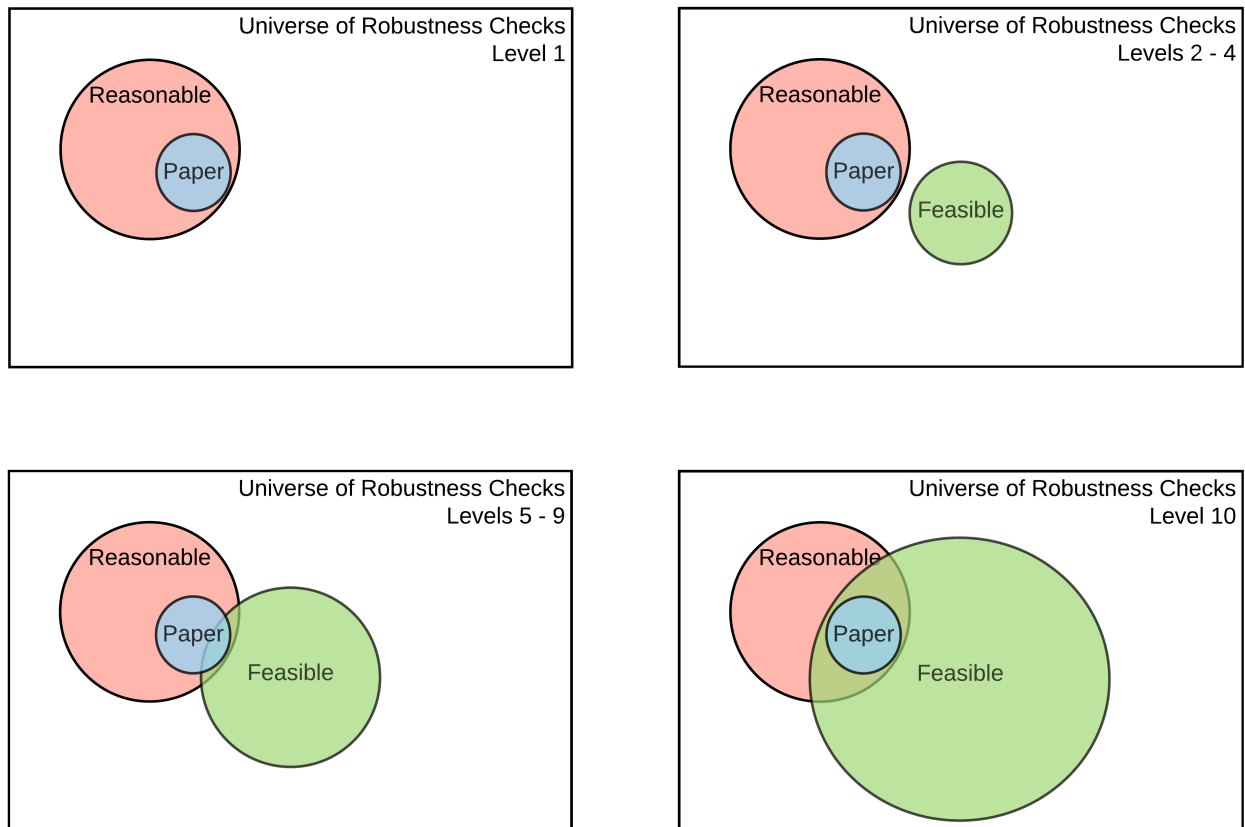


Figure 5.1: Universe of robustness tests and its elements

Here we suggest two types of contributions to robustness checks: (1) increasing the number of feasible robustness checks by identifying key analytical choices in code scripts and (2) justifying and testing reasonable specifications within the set of feasible checks. Both contributions should be recorded on the SSRP Platform and be linked to specific files in the reproduction package.

5.1 Feasible robustness checks: increasing the number of feasible specifications

Increasing the number of feasible robustness checks requires identifying the specific line(s) in the code scripts that execute an analytical choice. An advantage of this type of contribution is that you don't need to have an in-depth knowledge of the paper and its methodology to contribute. This allows you to potentially map several code files, achieve a broader understanding of the paper, and build on top of others' work. The disadvantage is that you are not expected to test and justify the reasonableness of alternative specifications.

Analytical choices can include those behind data cleaning and data analysis. Below are some proposed types for each category.

Analytical choices in data cleaning code

- Variable definition
- Data sub-setting
- Data re-shaping (merge, append, long/gather, wide/spread)
- Others (specify as "processing - other")

Analytical choices in analysis code

- Regression function (link function)
- Key parameters (tuning, tolerance parameters, and others)
- Controls
- Adjustment of standard errors
- Choice of weights
- Treatment of missing values
- Imputations
- Other (specify as "methods - other")

To record a specific analytical choice on the SSRP, please follow these steps:

1. Review a specific code file (e.g. `clean_merged_1_2.do`) and identify an analytical choice (e.g. `regress y x if gender == 1`).
2. Record the file name, line number, choice type, and choice value, as they appear in the original reproduction package. Sometimes the same analytical choice will be used more than one time in a given analysis (for example, a specific set of covariates is repeated across multiple regressions). In the **source** field, type "*original*" whenever the analytical choice is identified for the first time, and **file name-L+line number** each time the same analytical choice is applied thereafter (for example, if an analytical choice is identified for the first time in line #103 and for the second time in line #122 their respective values for the **source** field should be **original**, **code_01.do-L103**, and **code_01.do-L122** respectively). For each analytical choice recorded, add the specific choice used in the paper and, optionally, describe what alternatives could have been used in the **choice_range** column (e.g., "min, max" or "alt 1, alt 2, alt 3"). The resulting database would look like this (real reproduction example here):

entry_id	file_name	line_number	choice_type	choice_value	choice_range	source
1	code_01.do	73	data sub-setting	males	males, female	original
2	code_01.do	122	variable definition	income = wages + capital gains	wages, capital gains, gifts	code_01.do- L103
3	code_05.R	143	controls	age, income, education	age, income, education, region	original
...

The process of recording analytical choices can take a long time. Here a reproducer is not expected to identify all analytical choices. Ideally reproducers will focus on one script and map as much as possible within that script. Reproducers are encouraged to build on top of previous reproductions by exporting analytical choices of previous reproducers (by searching into previous reproductions and clicking “download table as csv” in the robustness section) and uploading them into their own reproduction (by clicking “upload csv”).

5.2 Justifying and testing reasonable robustness checks

Justifying and testing a specific robustness check involves identifying one or more feasible analytical choices, conducting a variation on them, and justifying its reasonableness. This approach’s advantage is that it allows for an in-depth inspection of a specific section of the paper. Its main limitation is that justifying sensibility and validity (and non-redundancy, to an extent) requires a deeper understanding of the paper’s topic and the methods. That may mean that undergraduate students or graduate students with only a surface-level interest in the paper (or limited time) may find it challenging to conduct this part of the reproduction.

When performing a specific robustness check, follow these steps:

1. Identify the set of analytical choices you will modify for a specific robustness test. Record the identifier(s) corresponding to the analytical choice(s) of interest (**entry_id**). For multiple identifiers, separate them using comma (e.g., **entry_id= 2, 5, 7**).
2. Propose a specific variation to these set of analytical choices.
3. Discuss whether you think these variations are sensible, specifically in the context of the claim tested (e.g., does it make sense to include or exclude low-income Hispanic people from the sample when assessing the impact of a large wave of new immigrants?).
4. Discuss how these variations could affect the validity of the results (e.g., likely effects on the omitted variable bias, measurement error, change in the Local Average Treatment Effects for the underlying population).
5. Confirm that this test is not redundant with other tests in the paper or in the robustness exercise.
6. Report the result of the robustness check (new estimate, standard error, and units), and discuss differences with the pre-specified estimates for this claim.

Discussing the reasonableness of a new robustness check could benefit from an open exchange of ideas. We recommend that reproducers create a new entry in the SSRP forum (remember to tag the DOI of the paper so others can find it) and post their robustness section in the forum to discuss with others (once public and in view only, you can share a link to any specific section of your reproduction).

Chapter 6

Concluding the Reproduction

Once you have completed each of the reproduction stages for all of your claims of interest, you will be ready to submit your work. For a reproduction to be considered complete and ready for submission, you must have **assessed at least one display item**.

Before submitting a reproduction, you will be able to modify your answers to any entry in the ACRE platform. After you hit “Submit”, however, you will not be able to modify your reproduction attempt any further. If you wish to modify your reproduction after submitting it, you will have to record *a new reproduction attempt* on the platform and link to the previously completed reproduction. **Tip:** the platform allows to clone entire reproductions up to the scoping stage. If you also want to bring in content from some of the labor-intensive parts in later stages (describing all the inputs, documenting analytical choices) you can download the csv file of previous forms and upload them to a new reproduction.

6.1 Outputs

A completed reproduction will consist of three different types of outputs:

1. Revised reproduction package – Deposit your revised version of the original reproduction package in a trusted repository, such as Dataverse, openICPSR, Figshare, Dryad, Zenodo, or the Open Science Framework. You should submit a revised reproduction package any time that you perform any type of improvement to the original reproduction package. This revised reproduction package is expected to be self-contained as it might be used by future reproducers for assessment and improvement. If your new reproduction package is larger than the capacity of your trusted repository (around 2GB), or it contains data that you do not have permission to share, remove the specific files from your reproduction package and add a reference to the original reproduction package. For example if your reproduction package contains a file that cannot be shared, locate it and leave a note in the readme file with its relative directory (e.g. `/data/large_confidential_file.csv`) and specific instructions on how to obtain it (see section 4.1.7 in Improvements).

Before submitting the reproduction, make sure that your revised reproduction package fulfills the following requirements:

- Has a digital object identifier (DOI). All of the trusted repositories referenced above generate DOIs. This is a stable, citable link that will allow others to easily find your work.
- Is titled following this convention: **revised reproduction package for - title of the paper - last name of reproducer - year when the reproduction was completed**
- If it is not self-contained, indicate in the readme file the DOI of the last reproduction package used for the reproduction (the original or a previously revised reproduction package), and the relative file location of the missing files (e.g. `/data/raw/large_file.csv`).

2. Shareable Reproduction: after submitting your reproduction, the platform will automatically generate a final report with its own DOI (different from the DOI used for the revised reproduction package). At the beginning of

your reproduction, the platform will also add a short summary indicating the main components of your work. You can share the entire report with your instructor or colleagues, or point them to a specific section (e.g., computation section in the Assessment stage).

3. Your data – You will be able to export several .csv files containing all of your table-form responses recorded as part of your reproduction attempt.

6.2 Visibility and data use

Before submitting your reproduction, we ask that you select one of three privacy models that determine what parts of your reproduction will be visible to other SSRP users upon publishing. We recommend using the *public, identified* model, which allows other users to see the entirety of your reproduction and credits you directly as its author. However, you may choose to publish your reproduction *anonymously* under either the *class anonymous* or *temporary anonymous* models, where other SSRP users can access unidentifiable data from your reproduction (e.g., numerical and categorical responses), but not your identity and identifiable data (free-form narrative answers such as notes, descriptions, summaries, and explanations). Please refer to the Terms of Use to learn more about each privacy model; however, here they are in a nutshell:

1. **Public, identified:** this is the default privacy setting. You are listed as the reproduction’s author, and other users can see the entire reproduction.
2. **Class anonymous:** allows you to conceal your identity and free-form narrative answers (e.g., notes, descriptions, summaries, explanations, but *not* one-sentence claim summaries as inputted in Q. 3.1.1 in *Scoping*) indefinitely if you conducted the reproduction as part of a class, project, or some other assignment for credit at the university/post-secondary level. Other users can see basic class identifiers (course name, year, institution university, instructor’s name, and email address), categorical responses, and your claim summaries.
3. **Temporary anonymous:** allows you to embargo your identity and free-form narrative answers (e.g., notes, descriptions, summaries, explanations, but *not* one-sentence claim summaries as inputted in Q. 3.1.1 in *Scoping*) for up to 4 years after submitting. Requires an annually recurring opt-out to maintain the embargo. At the expiration of the embargo period or if you fail to opt-out, privacy settings are updated to “public, identified”.

The Social Science Reproduction Platform will aggregate descriptive statistics from all submitted reproductions recorded to produce reproducibility metrics across disciplines, sub-disciplines, journals, and topical bodies of literature. If you choose to remain anonymous and are not a student using the SSRP as part of a course, your work may still be integrated into these descriptive statistics.

Chapter 7

Guidance for Instructors Supervising Reproduction Assignments

The Social Science Reproduction Platform (SSRP) is an openly-licensed platform that crowdsources and catalogs attempts to assess and improve the computational reproducibility of published social science research. Instructors can use the SSRP in combination with this Guide (called the “ACRe Guide”) to facilitate reproduction assignments (sometimes called “replications”) in applied social science courses at the graduate and undergraduate levels. Students can use these materials with little to no supervision, covering the following learning activities:

- Assessing and improving the reproducibility of published work
- Applying good coding and data management practices
- Recording reproduction results and improvements on the SSRP
- Engaging in constructive exchanges with authors
- Developing a deep understanding of common methods and computational techniques
- Creating real, citable scientific contributions.

In the process, students can reference the ACRe Guide as a step-by-step protocol for conducting reproductions and use the SSRP to record and share their work, and instructors can use it as a guide for supervising these assignments. Examples of completed reproductions are available [here](#).

Reproductions typically include five distinct stages (elaborated on at length in the chapters of this Guide):

1. *Paper selection*, where students select a paper and try to locate its reproduction package;
2. *Scoping*, where students define the scope of the reproduction by recording the claims, display items, and specifications to focus on;
3. *Assessment*, where students review and describe in detail the reproduction package, if available, and assess the current level of computational reproducibility of the selected display items;
4. *Improvement*, where students modify the content and/or the organization of the reproduction package to improve its reproducibility; and
5. *Robustness*, where students identify feasible robustness checks and/or assess the reasonableness of variations in analytical choices.

This chapter includes tips and resources for instructors interested in using the SSRP to teach reproducibility. We first identify typical use cases and suggest timelines for planning assignments. We then provide sample grading strategies and guidance in instances where the students would like to remain anonymous but still share their work on the SSRP.

Get started by signing up for a free SSRP account and reviewing this chapter to find tips based on your use case below.

7.1 Common use cases

7.1.1 Graduate-level assignment

(Estimated duration: 2 weeks to one semester)

Advanced master’s or PhD-level courses often feature assignments where students reproduce published papers to gain familiarity with fundamental concepts, research methods, and applications. In these courses, instructors can recommend students choose papers from specific sub-fields (e.g., labor economics) or journals, or curate a list of possible papers.

The length of such assignments can vary from a couple of weeks to a semester. For shorter timelines, instructors can ask students to conduct reproductions using only analysis data, focusing on just the main results. For longer timelines, such as semester-long assignments, students might reproduce entire papers from raw data. See section 7.2 below for a suggested distribution of effort across different lengths of assignments.

Graduate students should expect to spend a significant time at the *Robustness* stage. Therefore, before beginning the assignment, we recommend that students focus either on expanding the set of feasible robustness checks (see Chapter 5) or on testing and defending the reasonableness of a specific robustness check (which may contain more variations in more than one analytical choice).

7.1.2 Undergraduate thesis

(Estimated duration: 2 months to 1 year)

Students can use the platform to carry out detailed reproductions as part of an undergraduate thesis (see an example here). This type of assignment allows students to gain direct research experience while making meaningful contributions to the field.

Depending on the paper’s scope and the assignment’s learning objectives, such projects could last anywhere from a couple of months to an entire year. Students should demonstrate an understanding of the paper by identifying its main claim(s) at the *Scoping* stage and conducting a detailed assessment of the reproducibility of their associated display items at the *Assessment* stage. At the *Improvement* stage, students should try to improve the reproducibility of at least one display item. Depending on the scope of the paper, examples within reach of an undergraduate thesis might include finding and cleaning the raw data, translating code scripts into a different programming language (ideally open source), using dynamic documents, or making other improvements suggested by the instructor. Finally, at the *Robustness* stage, students should be able to demonstrate a high-level understanding of the conditions under which the paper’s estimates are valid.

We recommend breaking down the assignment into a “proposal” stage, where students scope and assess their papers, and an “execution” stage, where students carry out improvements and robustness checks.

7.1.3 Undergraduate course

(Estimated duration: 1 month to one semester)

Students in undergraduate courses should primarily focus on verifying whether reproductions packages “run” and recording the outcome on the SSRP. At the *Scoping* stage, undergraduate students can demonstrate a high-level understanding of the paper’s main claims. Students can identify the code and files necessary to reproduce their select display items at the *Assessment* stage and try to actually run the code. Finally, at the *Robustness* stage students can inspect the code scripts and identify potential lines of code with analytical choices, increasing the feasible set of robustness checks.

7.2 Paper selection and assignment timelines

The SSRP allows for flexibility in selecting papers and determining the scope of reproduction assignments in terms of the number of claims, display items, or even the required reproduction stages (at minimum, a complete

reproduction requires filling in *Scoping* and *Assessment*). For example, instructors can **curate a list of target papers** or **allow students to choose** from a distinct body of literature or a specific journal. For either strategy, a good starting point is first to reference the SSRP database of papers and reproductions, which allows users to search through submitted reproductions. This can help detect potential “no-go” papers for which other users were unable to find any reproduction materials (labeled as “abandoned papers”), though users are by no means barred from attempting to reproduce such cases. Reproducers can also look into the suggestions of future improvements (under the Improvements stage) and dedicate a reproduction exercise to a specific suggestion identified by in a previous reproduction.

If instructors are unable to curate a list of papers with readily available reproduction packages, we recommend allocating at least 3 additional weeks to try to locate the missing reproduction materials, potentially by contacting the original authors (see this chapter for guidance on constructive exchanges with original authors).

Instructors could select papers base on prior expectations of reproducibility, or based on proxies for impact papers on their field (e.g., citations). See the further guidance on paper selection in the Selecting a Paper chapter.

To plan the distribution of effort across different reproduction stages, we recommend consulting Table 1 below. The *Scoping* and *Assessment* stages will take approximately the same amount of time across timelines, but will depend on the students’ prior experience with reproductions and the SSRP. Larger differences emerge in the distribution of time for the last two main stages: Improvements and Robustness. For shorter exercises, we recommend avoiding any possible improvements to the raw data (or cleaning code), which may limit the number of possible robustness checks.

Table 1: Workload distribution across reproduction stages across different timelines

Stages of Reproduction	2 weeks (~10 days)		1 month (~20 days)		1 semester (~100 days)	
	analysis data	raw data	analysis data	raw data	analysis data	raw data
Scoping	10% (1 day)		10% (2 day)		10% (10 days)	
Assesment	35%		25%		15%	
Improvement	25%	0%	40%		20%	30%
Robustness	25%	5%	25%		25%	

7.3 Grading strategy

Once students have completed their reproductions, they can generate and share with instructors links to view-only reports of their work. Instructors can use these reports to grade and provide feedback on students’ work. Below we suggest a strategy for grading based on the use cases described above. For assignments that don’t contain all reproduction stages, we recommend distributing credits proportionally to the stages that were filled out.

We suggest the following grading categories:

- **Submission:** Full credit for timely submitted, non-blank submissions.
- **All key fields:** Full credit for filling out details such as paper information, reproduction package links, claims summary for at least one claim, table with claims estimates, inputs description, display item assessments (particularly the subjective assessment section), and assessment of paper-level reproducibility.
- **Paper summary:** Full credit for submitting a short summary of the paper that answers these questions.
- **Claims:** Full credit for presenting an accurate and clear summary of the paper’s main claims (see guidance here). To properly score this section, the instructor needs to have previous knowledge of the paper.

- **Assessment:** There are two kinds of contributions students can make to obtain full credit:
 - Assess *multiple claims*. Additional credit per each assessment could be assigned at a decreasing rate (e.g., in addition to the first assessment scored in “All key fields” give 50% of the credit for assessing a second claim, 30% for a third, and 20% for a fourth).
 - Conduct a *thorough assessment of a single claim*. This can be done by clearly identifying all of the inputs (data sources, analysis data, and code scripts) and providing detailed documentation of the reproduction attempt (e.g., duration of execution, errors detected, specific missing files, details about the required computational environment).
- **Improvement:** Students can obtain full credit through two types of contributions:
 - Implement improvements of reproducibility of *multiple claims* with a similar credit structure as for the assessment of multiple claims (see point above).
 - Implement *thorough improvements of a single claim*. The Improvements chapter describes several types of possible improvements. A key requirement is that students clearly document their improvements (see this section for suggestions on how to do this).
- **Future suggestions for improvement:** Instructors can also assign credit for specific suggestions for improvements recorded for future reproducers to attempt. Full credit should be assigned when the recommendations are concrete, feasible, and demonstrate that the student performed an in-depth review of the code and data.
- **Revised reproduction package:** Full credit for students who post a revised reproduction package in a trusted repository. Instructors can verify its existence by clicking on the link provided on the reproduction summary page. If the student used version control software as part of the revised reproduction package, the instructor can easily verify the modifications relative to the original reproduction package by checking the commits history.
- **Robustness:** Students can obtain full credit for one of two strategies (or a combination of them):
 - Identify a large number of original analytical choices in the code files. The definition of “large” is up to the Instructor, but we suggest at least 20 original analytical choices (see this section for a distinction between original and non-original analytical choices).
 - Conduct a robustness test by modifying one or more analytical choices identified in the paper. Full credit is assigned for a clear and accurate discussion of the results and their reasonableness.

	Score (example)	Weights (suggested)
Submitted	100	40%
All important fields	80	10%
Summary	100	5%
Clear claims	90	5%
Assessment	100	10%
Improvements	50	10%
Revise reproduction package	100	5%
Specific suggestions for improvements	0	5%
Robustness	100	10%
Final Score	87.5	

7.4 Facilitating *class anonymous* reproductions

The SSRP allows users to select from three distinct privacy models when submitting their work, including *public-identified*, *temporary anonymous*, and *class anonymous* (explained in more detail here). This section includes specific guidance for instructors facilitating reproductions that will be published using the *class anonymous* privacy model, which allows users to conceal their identities and free-form narrative answers (e.g., notes, descriptions, summaries, explanations) indefinitely. Instead, reproductions published under the class anonymous model contain only categorical responses and basic class identifiers, including course name, year, institution, instructor’s name, and instructor’s email address.

Reproducers looking to publish under the *class anonymous* model will be unable to maintain their anonymity if they share their revised reproduction packages as part of their reproductions. This is because trusted data repositories and GitHub require users to create accounts in order to post their content and keep records of updates. We advise instructors to create generic repositories for their class (e.g., “Reproductions for Econ 270 - Fall 2021”) and deposit each anonymous reproduction in self-contained folders.

For example, suppose a class has two students who choose the anonymous option. In that case, they should share (privately) their revised reproduction package with their instructor, and the instructor should post them using the following structure:

```
Reproductions for [Course Name, Semester]
  Revised reproduction package for Smith et al. (2019)
  Revised reproduction package for Perez et al. (2019)
```

Note: When anonymizing revised reproduction packages, we advise that instructors and students avoid using version control software. Even if the current materials are correctly de-identified, the version history may still contain identifiers, such as GitHub usernames.

Chapter 8

Guidance for Constructive Communication Between Reproducers and Original Authors

This chapter contains guidance for constructive and respectful communication between reproducers and original authors. Exchanges that contain charged or adversarial language can damage professional relationships and hamper scientific progress. Janz and Freese (2019) articulate two important steps reproducers can take to ensure that their interactions with original authors are constructive. We summarize and build on this approach below and encourage you to follow this guidance. Remember the **golden rule of reproductions** (and replications): *treat others and their work, as you would like others to treat you and your work!*

1. Carefully and transparently plan your study.

- a. Clearly state that you are conducting a reproduction of their original work.
- b. Explain why you have chosen this study.
- c. If you are not able to reproduce the results explain how “far” your results must deviate from the original work before claiming that the study could not be reproduced. Engage deeply with the substantive literature to ensure that your interpretation of differences between the original and reproduction is thorough and acceptable to other authors in the field.

2. Use professional and sensitive language. Discuss potential discrepancies between your work and the original paper, just like you might do for your own work.

- a. Avoid binary judgments and statements like “failed to reproduce.” Instead, state clearly which results reproduced and which did not (e.g., “we successfully reproduced X, but failed to reproduce Y”) unless you uncover apparent scientific misconduct (e.g., see Broockman, Kalla and Aronow, 2015).
- b. Talk about *the study*, not *the author* to avoid making it personal. Make clear what the positive contribution of the original article is. Consider sending a copy of your reproduction report to the original authors.
- c. Discuss what your reproduction contributes to the literature, and refrain from claiming to give the final answer to the question.
- d. For papers published five or more years ago, be mindful that norms for reproducibility have evolved since then.
- e. Remember, *the goal is not to criticize previous work or hunt for errors, but to move the literature forward!*

To help you put these recommendations into practice, we developed template language for common scenarios that reproducers and authors may encounter in their interactions. While we hope you find these useful, note that they

are *only recommendations*, and you are welcome to modify them based on your project’s context and needs. Feel free to contact us if you need more guidance or would like to provide feedback on these materials.

8.1 For reproducers contacting the authors of the original study

Consider the following *before* you contact the original authors:

1. Search the SSRP for reproduction materials or records of past interactions with the authors of a given paper. For each reproduction, consult the “Author inquiries” column to see whether authors have indicated whether they are available for further inquiries.
2. Carefully read all footnotes, appendices, tables, captions, etc., to learn if, how, and where reproduction materials are provided. Follow this Data and Code Guidance to determine whether you have everything before you start. A few things to consider:
 - A *Readme* file, if available, would be a good place to start.
 - Check whether there are any restrictions on accessing the data or code, and whether there are instructions on how to access these files for the purpose of reproduction.
3. If a reproduction package is not readily available in the location where the article is published (e.g., the journal website), check the authors’ websites, Dataverse profiles, or other relevant archives and/or data repositories like the ICPSR Publications Related Archive.
4. If steps 1 and 2 don’t yield anything, contact the corresponding author (copying the co-authors, if any), consolidating your requests into as few emails as possible. In your email, make sure to include the following details:
 - Basic information about the paper to reproduce (include title, version, date, and a DOI (or just a URL));
 - Context for the reproduction (as part of a class exercise, thesis, personal project, etc.) and a note that the outcome will be recorded on the Social Science Reproduction Platform(SSRP);
 - Items from the reproduction package that are missing, as well as locations where you had (unsuccessfully) searched for them;
 - Your use plan: Will the materials be used exclusively for this project? Ask for permission to share the data publicly.
 - Right to consultation and results: Will you share the outcome of the reproduction with the original authors?
 - A deadline to respond (we suggest at least two weeks).
5. Follow up if you don’t get a response within two weeks (or whatever deadline you set), and include any details or clarifications that were left out in your first email.
6. Record the outcome of your interaction with the original author at the *Select a Paper* stage on the SSRP. You can qualify the outcome as one of the following:
 - A *complete* reproduction package was provided;
 - An *incomplete* reproduction package was provided. You can also select one of the following reasons:
 - Data is of sensitive, confidential, or proprietary character and cannot be shared;
 - Data is of sensitive, confidential, or proprietary character, but access instructions were provided.
 - The author *declined* to share the reproduction package; or
 - The author *did not respond* (including after a reminder was sent) within four weeks after the initial request.

8.1.1 Contacting the original author(s) when there is no reproduction package

Template email:

Subject: Reproduction package for ["Title of the paper"]

Dear [Title (e.g., "Dr.") Last name of Corresponding Author],

I am contacting you to request a reproduction package for your paper titled [Title] which was published in [Journal] in [year] (vol [volume], no. [no.]), [link]. A reproduction package contains (raw and/or analytic) data, code, and other documentation that makes it possible to reproduce the paper. Would you be able to share any of these items?

I am a [graduate student/postdoc/other position] at [Institution], and I would like to reproduce the results, tables, and other figures using the reproduction materials mentioned above. I have chosen this paper because [add context for why you want to reproduce this particular paper using neutral language (e.g., "This is a seminal paper in my field"), avoiding any statements that would put the respondent on the defensive]. Unfortunately, I was not able to locate any of these materials on the journal website, Dataverse [or other data and code repositories], or your website.

I will record the result of my reproduction attempt on the Social Science Reproduction Platform (SSRP), an open-source platform for systematically conducting and recording reproductions. With your permission, I will also record the materials you share with me, which would allow access for other reproducers and avoid repeated requests directed to you. Please let me know if there are any legal or ethical restrictions that apply to any of the reproduction materials so that I can take that into consideration during this exercise.

In addition to your response above, would you be available to respond to future (non-repetitive) inquiries from me or other SSRP users? Though your cooperation with my and/or future requests would be extremely helpful, please note that you are *not required to respond*.

Since I am required to complete this project by [date], I would appreciate your response by [deadline].

Let me know if you have any questions. Please also feel free to contact my supervisor/instructor [Name (email)] for further details on this exercise. Thank you in advance for your help!

Best regards,
[Reproducer]

8.1.2 Follow-up if the author only provides the appendix

If the original author asks you to reproduce the results using the appendix, you are not obligated to undertake that effort. Should you choose to do so, you can send a follow-up email as follows:

Dear [Title (e.g., "Dr.") Last name of Corresponding Author],

Thank you for your response. The purpose of my reproduction is to assess and improve computational reproducibility using the original data and code. Your appendix was very helpful, and I have attempted to use it to reproduce your results. [Describe initial efforts to reproduce the results.].

To help me advance with the reproduction, I hope you can share the source code and provide guidance on the following: [List out any unclear steps, data sources, or otherwise missing components. Use bullet points if more than two].

Thank you in advance for your help! Once completed, I will make the reproduction package publicly available to be used in future SSRP reproductions. Please let me know if you have any questions.

Best regards,
[Reproducer]

8.1.3 Contacting the original author(s) to request specific missing items of a reproduction package

Template email:

Subject: Reproduction materials for ["Title of the paper"]

Dear [Title (e.g., "Dr.") Last name of Corresponding Author],

I am contacting you regarding reproduction materials for your paper titled [Title] was published in [Journal] in [year] (vol [volume], no. [no.]), [link].

I am a [graduate student/postdoc/other position] at [Institution], and I'm working on reproducing this paper as part of a class assignment. [Add context for why you want to reproduce this particular paper using neutral language (e.g., "This is a seminal paper in my field"), avoiding any statements that would put the respondent on the defensive].

To help me reproduce the paper in full, I hope that you can share the following items: [list items missing from reproduction package, preferably bulleted if more than one (e.g., raw/analytic data, code, protocols for conducting the experiment, etc.)]. I have already searched [locations where you searched for items, with links provided], but I could not locate the items. Unless you tell me otherwise, I will make the reproduction package publicly available to be used in future reproductions. Let me know if any legal or ethical restrictions apply to any of the reproduction materials to consider during this exercise.

Note that I will record the outcome of my reproduction on the Social Science Reproduction Platform(SSRP), an open-source platform for systematically conducting and recording reproductions. Let me know if you would like me to share the outcome of my reproduction with you, and whether you are interested in providing a response.

Since I am required to complete this project by [date], I would appreciate your response by [deadline].

Let me know if you have any questions. Please also feel free to contact my supervisor/instructor [Name (email)] for further details on this exercise. Thank you in advance for your help!

Best regards,
[Reproducer]

8.1.4 Asking for additional guidance when some materials have been shared

Note: Even when a corresponding author has shared a reproduction package, you may still face challenges in interpreting or executing the materials. That shouldn't discourage you from asking the corresponding author to provide clarifications or to share the missing materials. As in the previous scenario described above, demonstrate that you've made an honest effort to reproduce the work using the available resources and try to consolidate your requests in as few emails as possible.

Template email:

Subject: Clarification for reproduction materials for ["Title of the paper"]

Dear [Title (e.g., "Dr.") Last name of Corresponding Author],

Thank you for sharing your materials. They have been immensely helpful.

Unfortunately, I ran into a few issues as I delved into the reproduction, and I think your guidance would be helpful to resolve them. [Describe the issues and how you have tried to resolve them. Describe whatever files or parts of the data or code are missing. Refer to examples 1 and 2 below for more details].

Thank you in advance for your help.

Best regards,
[Reproducer]

1: An example of a well-described issues:

Specifically, I am attempting to reproduce Display Item X (e.g., table 1, figure 3). I found that the following components are required to reproduce Display Item X:

```

Display Item X
  [code] formatting_table1.R
    display_itemx_part1.txt
  /      [code] output_table1.do
  /      [data] analysis_data01.csv
  /      [code] data_cleaning01.R*
  /      [data] UNKNOWN
    display_itemx_part2.txt
      [code] output_table2.do
      [data] analysis_data02.csv
      [code] data_cleaning02.R
      [data] admin_01raw.csv*

```

I have marked with an asterisk the items that I could not find in the reproduction materials: **data_cleaning01.R** and **admin_01raw.csv**. After accessing these files, I will also identify the name of the raw data set required to obtain output1_part1.txt. This is to let you know that I may need to contact you again if I cannot find this file (labeled as UNKNOWN above) in the reproduction materials.

I understand that this request will require some work for you; I will publish the materials on the Social Science Reproduction Platform, which should help avoid repeated requests in the future.

2. An example of a poorly-described issue:

Your paper does not reproduce. I have tried for several hours now, and can't get the DO files to run. Could you please share all the missing reproduction materials? Data and code sharing are basic principles of open science, so I am confident that you will do the right thing.

8.1.5 Response when the original author has declined to share due to *undisclosed reasons*

Note: You can also use this template if a corresponding author has not submitted a response after two or more follow-up emails.

Template email:

Subject: Re: Reproduction materials for ["Title of the paper"]

Dear [Title (e.g., "Dr.") Last name of Corresponding Author],

Thank you for considering my request. I will try to reproduce the paper using the available materials and will record the missing items accordingly on the Social Science Reproduction Platform (SSRP). I will also post my assessment of the reproducibility of the paper in its current form based on the SSRP reproducibility scale.

Let me know if you have any questions.

Best regards,
[Reproducer]

8.1.6 Response when the original author has declined to share due to legal or ethical restrictions of the data

Template email:

Subject: Re: Reproduction materials for ["Title of the paper"]

Dear [Title (e.g., "Dr.") Last name of Corresponding Author],

Thank you for your response and for clarifying the terms of use for the reproduction materials.

Though I understand you are unable to share the raw data, there may be alternative steps you can take that would help me improve the reproducibility of your paper. These include:

1. Sharing the analytic version of the data (the version of the dataset that was used for analysis in the final version of your paper);
2. Providing a public description of the steps other researchers can follow to request access to the raw data or materials, including an estimate of the costs and the duration of the process. You can find examples of data availability statements for proprietary or restricted-access data [here](#); and
3. Providing access to all data and materials for which the constraints do not apply.

Based on my assessment, your paper would currently rank at [level X] on the SSRP reproducibility scale. However, *this score can be easily improved*. Being able to provide analytic data would elevate the reproducibility of your paper to [level Y]. Providing public instructions on how other parties can access the data would further elevate its reproducibility to [level Z].

I would be happy to help if you are interested in taking any of the steps I outlined above.

Thank you for your help!

Best regards,
[Reproducer]

8.1.7 Contacting the original author to share the results of your reproduction exercise

Note: Reporting the results of reproductions can be the most contentious part of the process, particularly when the reproducer is unable to fully reproduce the paper or finds significant deviations from the original work. However, if the reproduction can correctly identify the sources of such deviations, it may be viewed as an improved version of the original work.

Regardless of the outcome of the reproduction exercise, the guidance from the introduction of this chapter still stands: *reproduce the work of others as you would like for others to reproduce yours*, and make sure that is reflected in how you discuss any discrepancies between your and the original work.

Template email:

Subject: Reproducibility Assessment of ["Title of the paper"]

Dear [Title (e.g., "Dr.") Last name of Corresponding Author],

Thank you for your support throughout my project as I worked to verify and advance the reproducibility of [Paper]. I'm writing now to share the results of my project and to invite your feedback.

The results of each step of my reproduction include i) Assessment, ii) Improvements, iii) Robustness Checks, (and iv) Extensions, if applicable).

‘[Include the following items in the body of your email:]’

I look forward to your questions, comments, and suggestions for my work. As discussed previously, I will record the outcomes of my reproduction, along with the improvements, on the Social Science Reproduction Platform.

Best regards,
[Reproducer]

8.1.8 Responding to hostile responses from original authors

Note: Planning your study carefully and transparently and using professional and sensitive language are the best ways to ensure that the interaction will be beneficial to both you and the original author. However, unpleasant interactions may happen despite your best efforts and can range anywhere from dismissive comments to bullying, discrimination, and harassment. Find guidance at the end of this chapter on how to deal with instances of bullying, harassment, or discrimination.

8.1.8.1 Dismissive comments

In cases of dismissive comments, the best course of action may be to simply thank the author for their response and continue with the reproduction.

Template email:

Subject: Re: Reproduction materials for ["Title of the paper"]

Dear [Title (e.g., "Dr.") Last name of Corresponding Author],

Thank you for your response. I will work to reproduce your paper using the available materials and will record my results accordingly on the Social Science Reproduction Platform. I will also post my assessment of the reproducibility of the paper in its current form based on the SSRP reproducibility scale.

Let me know if you have any questions.

Best regards,
[Reproducer]

8.2 For original authors responding to requests from reproducers

This section contains guidance for authors whose work is involved in reproductions on the Social Science Reproduction Platform. We present language that may help various scenarios in which authors find themselves when interacting with reproducers. Though every interaction between authors and reproducers takes place is distinct and may carry its unique challenges, the guiding principle of this chapter always applies: “Treat others and their work as you would like others to treat you and your work!” We hope that these resources will facilitate more efficient and constructive exchanges between the parties involved. Let us know if you need guidance in other scenarios!

8.2.1 Responding to a repeated request that has been addressed in an earlier interaction

Dear [Reproducer],

Thank you for your interest in my work. I have been contacted about this issue by another SSRP reproducer before and provided a response, which I suspect may be already recorded on the SSRP. I’m copying my original response below for your reference. You may find further guidance in the readme file in the reproduction package.

If there are no prior records of these issues on the SSRP, please record the enclosed response [and materials]. This will also help avoid the duplication of effort on the part of others who may be interested in reproducing this work.

Good luck with the remainder of your project and thank you in advance for your cooperation!

Best regards, [Author]

8.2.2 Acknowledging that the author no longer has access to certain part(s) of the reproduction package

Dear [Reproducer],

Thank you for reviewing my work closely. I wish I could be of more help, but unfortunately I no longer have access to the requested materials due to [briefly describe the circumstances that prevent you from providing the materials].

While I recognize that the current standards in the discipline have moved towards computational reproducibility, note that this paper was written when different standards applied. Please feel free to evaluate the paper as is and propose any improvements wherever possible.

I look forward to working with you to address this and improve the overall reproducibility of the paper.

Best regards, [Author]

8.2.3 Acknowledging that some material is still embargoed for future research

Dear [Reproducer],

Thank you for your interest in my work. The data/materials/program that you reference are currently not publicly available because they are embargoed until [embargo period].

[Depending on the restrictions that apply to the reproduction package, consider alternatives to sharing the reproduction materials in full. These include: 1. Sharing the analytic version of the data (the version of the dataset that was used for analysis in the final version of your paper); 2. Providing a public description of the steps that other researchers can follow to request access to the raw data or materials, including an estimate of the costs and the duration of the process. Find examples of data availability statements for proprietary or restricted-access data [here] (<https://social-science-> and 3. Providing access to all data and materials for which the constraints do not apply.)

I hope you find this helpful. Please feel free to contact me if you have any further questions.

Best regards, [Author]

8.2.4 Responding to incomplete/unclear requests

Dear [Reproducer],

Thank you for your interest in my work. I would be happy to assist you and other reproducers to assess and improve the reproducibility of this paper.

To help me give more concrete guidance, I'd appreciate if you could provide a more specific description of the items that you need from me. You can find helpful information and resources in Chapter 6 of the Guide, >specifically here [based on the context, you may need to point >the reproducer to a >different scenario and/or provide further information].

Feel free to contact me if you have any further questions. Thank you for your cooperation.

Best regards, [Author]

8.3 Harassment and/or discrimination

The American Economic Association (AEA) and other academic societies have strict policies against harassment and discrimination. Here are some of the behaviors that the AEA Policy on Harassment and Discrimination has listed as unacceptable and could emerge in a hostile exchange regarding a reproduction:

- Intentionally intimidating, threatening, harassing, or abusive actions or remarks (both spoken and in other media)

- Prejudicial actions or comments that undermine the principles of equal opportunity, fair treatment, or free academic exchange
- Deliberate intimidation, stalking, or following
- Real or implied threat of physical harm.

Here are a some steps you can take if you believe you have experienced bullying, discrimination or harassment:

- **File a complaint with the AEA Ombudsperson.** Any AEA member can file a complaint. You can also join the AEA solely to file a report. The person about whom you are making the complaint need not be an AEA member. A non-AEA member can also file a report if the act of harassment or discrimination was committed by an AEA member or in the context of an AEA-sponsored activity. Learn more about the process [here](#).
- **File a report with your institution's office for the prevention of harassment & discrimination.** US-based institutions have internal mechanisms that allow students and faculty to seek support in cases of discrimination and harassment based on race, color, national origin, gender, age, or sexual orientation/identity, including allegations of sexual harassment and sexual violence. Formal titles of this office vary across institutions, but common names include "Office for the Prevention of Harassment and Discrimination" (in institutions that are part of the University of California system), "Office of Equity and Title IX," etc.
- **Contact your institution's Ombudsperson/Ombuds Office.** If you believe that you have experienced academic bullying or other forms of disrespectful behavior that fall outside the scope of harassment and/or discrimination as described above, you should know that university ombuds officers offer a confidential, impartial resource to discuss your concerns and learn about potential next steps available in your case.
- **Access mental health services at your institution.** While no amount of bullying, discrimination, or harassment is acceptable or the fault of the victim, these unfortunately still occur and can take a toll on victims' mental health. Many universities offer short-term Counseling & Psychological Services (CAPS) for academic, career, and personal issues.
- **Ask for support from your academic supervisor.** If you are unsure on how to proceed, consult your academic supervisor on whether continuing the reproduction is appropriate.

Chapter 9

Examples of Reproduction Trees

A diagram generated by the Diagram Builder which represents all the available data and code on behind a specific display item. The tree is meant to represent the entire computational workflow behind a result from the paper. It allows reproducers to trace a display item to its primary sources. It can also be used to guide users of the reproduction package and/or to identify missing components for a complete reproduction.

A reproduction tree is complete when it is possible to connect its output (a given display item) with all of its inputs down to the raw data. A reproduction is incomplete when it is not possible to connect all the inputs to the resulting display item. Paraphrasing the author Leo Tolstoy, *complete workflows are all alike; every incomplete computational workflow is incomplete in its own way*. This chapter presents a few examples of reproduction trees, focusing particularly on the many possible ways in which a tree could be incomplete. If you have a reproduction tree that contains an instructive example please contribute to this chapter (via a pull request or emailing your reproduction tree to ACRE@berkeley.edu.)

9.1 Stylized examples

9.1.1 Complete reproduction tree

Below is an example of output from the Diagram Builder for a display item that can be fully constructed using the files contained in the reproduction package. The diagram displays files as outputs and inputs to code scripts all the way down to the raw data.

```
table 1
  [code] formatting_table1.R
    output1_part1.txt
    |   [code] output_table1.do
    |       [data] analysis_data01.csv
    |       [code] data_cleaning01.R
    |       [data] survey_01raw.csv
    output1_part2.txt
      [code] output_table2.do
      [data] analysis_data02.csv
      [code] data_cleaning02.R
      [data] admin_01raw.csv
```

9.1.2 Incomplete reproduction tree

9.1.2.1 Raw data and analytic data are available, but cleaning code is missing.

Below is an example of output from the Diagram Builder for a display item that is missing some of the code needed to generate it from the raw data. There are two reasons to suspect that this workflow is incomplete: (i) there is no clear data cleaning step (only analysis that generates output and formatting), and (ii) there are unused files that are likely to be raw data. None of these reasons can confirm unequivocally that the tree is incomplete, but a reproducer familiar with the paper and its data sources could use the tree to certify its (in)completeness and request missing files.

```

table 1
  [code] formatting_table1.R
    output1_part1.txt
  |   [code] output_table1.do
  |   [data] analysis_data01.csv
    output1_part2.txt
      [code] output_table2.do
      [data] analysis_data02.csv

Unused files:
- survey_01raw.csv
- admin_01raw.csv

```

Reproducers are asked to speculate on where the missing files might go, and hence propose how a complete tree might look like (where possible). For this example, we have assumed there are missing code scripts that at some point take in `survey_01raw.csv` and `admin_01raw.csv`, and eventually output `analysis_data01.csv` and `analysis_data02.csv`, though this requires the reproducer's discretion.

```

table 1
  [code] formatting_table1.R
    output1_part1.txt
  |   [code] output_table1.do
  |   [data] analysis_data01.csv
  |   [code] MISSING FILE(S)
  |   [data] survey_01raw.csv
    output1_part2.txt
      [code] output_table2.do
      [data] analysis_data02.csv
      [code] MISSING FILE(S)
      [data] admin_01raw.csv

```

9.1.3 Unused data sources

It is possible that not all data included in a replication package are actually used in code scripts in the reproduction package. This would be the case if, for example, the raw data and analysis data are included, but not the script that generates the analysis data. As a concrete example, consider what the original diagram above would look like if the only code included in the reproduction package were `analysis.R`:

```

table1.tex
  |___[code] analysis.R
  |___analysis_data.dta

Unused data sources:
raw_1.dta

```

```

raw_2.dta
raw_3.dta
raw_4.dta

Unused analysis data:
cleaned_1.dta
cleaned_2.dta
cleaned_3.dta
cleaned_4.dta
merged_1_2.dta
merged_3_4.dta
cleaned_1_2.dta
cleaned_3_4.dta

```

In this case, there are many data files that were listed in the raw data and analytic data spreadsheets that are not used by any code script in the replication package.

9.1.4 Final outputs is not a display item

9.2 Examples from real reproduction attempts

9.2.1 Possibly missing code for producing a display item¹.

This reproduction diagram fragment likely shows a missing piece of code. In a complete reproduction package there are no unused files and all final outputs are display items. `cps2018.dta` and `cps_march2017.dta` are included as inputs in the reproduction kit, but never used to make a display item, i.e., there is no piece of code that is listed as using them as inputs. Likewise, `PublicSalary.dta` is listed as a final output, meaning it, too, is not used to make a display item since it is not listed as an input for any code script. Perhaps `Paper2_PoExitDataset.dta` is made by a missing code script that takes `PublicSalary.dta`, `cps2018.dta`, and `cps_march2017.dta` as inputs? It may be the case that the only way to find out for sure is to contact the study author(s), in which case, such a diagram can be used to help them identify any missing files.

```

PublicSalary.dta
|__MakeData2.do
|    |__Paper2_ProviderDataset.dta
|    |__Paper2_SSPDataset_wIRT_final.dta
|    |__PublicFacilitySurvey_clean.dta

Table1.xml
|__Table1.do
|    |__ProviderData.dta
|    |    |__MakeData4.do
|    |    |    |__Paper2_PoExitDataset.dta
|    |    |    |__Paper2_ProviderDataset.dta
|    |    |    |__Paper2_SSPDataset_wIRT_final.dta
|    |__VillageDataset.dta
|    |    |__MakeData8.do
|    |    |    |__Paper2_HouseholdDataset1.dta
|    |    |    |__Paper2_VillageDataset.dta
|    |__HouseholdDataset.dta
|    |    |__MakeData8.do
|    |    |    |__Paper2_HouseholdDataset1.dta
|    |    |    |__Paper2_VillageDataset.dta

```

¹This is from a reproduction attempt conducted as part of a UC Berkeley Development Economics course

```

Unusued data sources:
cps2018.dta
cps_march2017.dta

```

9.2.2 Long, complicated tree².

This diagram shows that the production of a given display item may be very complicated, highlighting the usefulness of the Diagram Builder as a visualization tool. In reproducing such a complicated display item, it can be useful to have such a diagram to determine, for example, the order in which code scripts should be run, what files might depend on a faulty code script, or which files are necessary to keep if the goal is to only produce the specific display item.

9.2.3 Example of completing reproduction tree

In many cases, some of the components of the workflow will not be easily identifiable (or missing) in the reproduction package. Here we present a more complex example than the one presented in the Assessment chapter. The Diagram Builder will return a partial reproduction tree diagram. For example, if the files `merge_1_2.do`, `merge_3_4.do`, and `final_merge.do` are missing from the previous diagram, the Diagram Builder will produce the following diagram:

```

cleaned_3.dta
  [code] clean_raw_3.py
        raw_3.dta

table1.tex
  [code] analysis.R
        analysis_data.dta

cleaned_3_4.dta
  [code] clean_merged_3_4.do
        merged_3_4.dta

cleaned_1.dta
  [code] clean_raw_1.py
        raw_1.dta

cleaned_2.dta
  [code] clean_raw_2.py
        raw_2.dta

cleaned_4.dta
  [code] clean_raw_4.py
        raw_4.dta

cleaned_1_2.dta
  [code] clean_merged_1_2.do
        merged_1_2.dta
Unused data sources: None.

```

In this case, you can still manually combine this partial information with your knowledge from the paper and own judgement to produce a “candidate” tree diagram (which might lead to different reproducers recreating different diagrams). This may look like the following:

²This is from a reproduction attempt conducted as part of a UC Berkeley Development Economics course

```
table1.tex
  [code] analysis.R
    analysis_data.dta
      MISSSING_CODE_FILE_3
        cleaned_3_4.dta
          | [code] clean_merged_3_4.do
          | merged_3_4.dta
          | MISSSING_CODE_FILE_2
          | cleaned_3.dta
          | | [code] clean_raw_3.py
          | | raw_3.dta
          | | cleaned_4.dta
          | | [code] clean_raw_4.py
          | | raw_4.dta
        cleaned_1_2.dta
          [code] clean_merged_1_2.do
          merged_1_2.dta
            MISSSING_CODE_FILE_1
              cleaned_1.dta
                | [code] clean_raw_1.py
                | raw_1.dta
                |
              cleaned_2.dta
                [code] clean_raw_2.py
                raw_2.dta
```

To leave a record of the reconstructed diagrams, you will have to amend the input spreadsheets using placeholders for the missing components. In the example above, you should add the following entries to the code description spreadsheet:

Table 9.1: Adding rows to code spreadsheet					
file_name	location	inputs	outputs	description	primary_type
...
missing_file1	unknown	cleaned_1.dta, cleaned_2.dta	merged_1_2.dta	missing code	unknown
missing_file2	unknown	cleaned_3.dta, cleaned_4.dta	merged_3_4.dta	missing code	unknown
missing_file3	unknown	merged_3_4.dta, analysis_data.dta merged_1_2.dta		missing code	unknown

Chapter 10

Contributions

10.1 Contributing feedback on this Guide

We welcome feedback and direct contributions from all readers in all parts and aspects of the ACRE Guide. If you wish to provide feedback on specific chapters or sections, click the “edit” icon at the top of this page (this will prompt you to sign into or create a GitHub account), after which you’ll be able to suggest changes directly into the text. Please submit your suggestions using the “create a new branch and start a pull request” option and provide a summary of the changes you’ve proposed in the description of the pull request. The ACRE project team will review all suggested changes and decide whether to “push” them to this Guide document or not.

To report issues or provide feedback on the Social Science Reproduction Platform, please submit to the Help category in the SSPP Forum or contact ACRE@berkeley.edu.

Major contributions to this Guide will be acknowledged below according to the Contributor Roles Taxonomy (CRediT). Major contributions are defined as any pushed revisions to the Guide language or source code beyond corrections of spelling and grammar.

10.2 List of Contributors: Guidelines content and source code:

(in alphabetical order)

- Aleksandar Bogdanoski – Funding acquisition, Project administration, Writing (original draft), Writing (reviewing and editing)
- Carson Christiano – Funding acquisition, Project administration, Writing (reviewing and editing)
- Joel Ferguson – Writing (original draft), Writing (reviewing and editing)
- Fernando Hoces de la Guardia – Conceptualization, Funding acquisition, Writing (original draft), Writing (reviewing and editing)
- Katherine Hoerberling – Funding acquisition, Project administration, Writing (original draft), Writing (reviewing and editing)
- Edward Miguel – Conceptualization, Funding acquisition, Supervision
- Emma Ng – Visualization, Writing (original draft), Writing (reviewing and editing)
- Lars Vilhuber – Conceptualization, Funding acquisition, Supervision

The individuals below have contributed to the ACRE GitHub repository:

@abogdanoski, @albertchae, Emma Ng (@em-ng21), Fernando Hoces de la Guardia (@fhoces), Joel Ferguson (@joelferg), Katie Hoerberling (@khoerberling), Michael Weiss (@mweiss), max zhang (@myzhang01), Peter Zhang (@petezh), Tarangana Thapa (@taranganathapa)

10.3 Suggested citation format

These follow the Chicago style citation format. Brackets [] indicate where you'll need to input more specific information.

- When citing the ACRE Guide: Berkeley Initiative for Transparency in the Social Sciences. 2020. “Guide for Advancing Computational Reproducibility in the Social Sciences.” [Date accessed (Day, Month, Year)]. <https://bitss.github.io/ACRE/>.
- When citing reproductions accessed on the Social Science Reproduction Platform: [Reproducer Last name], [Reproducer First name], [Last, First], and [Last, First]. [Year]. “Reproduction of [Title of original paper].” *Social Science Reproduction Platform*. doi: [doi]. [DOI or link to original paper].

10.4 Acknowledgments

Support for the development of this Guide was provided by Arnold Ventures.

Chapter 11

Definitions

- **Abandoned paper:** A paper that has been considered for reproduction, but the reproducer decided not to move forward with the analysis due to failure to locate a reproduction package. [Learn more here.](#)
- **Analysis code:** A script associated primarily with analysis. Most of its content is dedicated to actions like running regressions, running hypothesis tests, computing standard errors, and imputing missing values.
- **Analytic data:** Data used as the final input in a workflow to produce a statistic (display item) in the paper (including appendices).
- **Candidate paper:** A paper that is currently been considered for reproduction. The reproducer can move forward to scope the paper or abandon the paper if there is no reproduction package. [Learn more here.](#)
- **(Research) Claim:** According to a definition provided by the repliCATS project, a research claim is a single major finding from a published study, as well as details of the methods and results that support this finding. A paper can include more than one claim. Claims are usually described in the paper’s abstract; however, sometimes, the claim described in the abstract does not match the claim that is tested in the paper. In the SSRP framework, different claims in a paper may be tested using different methodologies, and their results may be presented in one or more display items, such as tables and figures (figure 0.1 illustrates this idea). There are different types of claims, including:
 - **Causal claim:** An assertion that invokes causal relationships between variables. A paper may estimate the effect of X on Y for population P , using method F . E.g., “This paper investigates the impact of bicycle provision on secondary school enrollment among young women in Bihar/India, using a Difference in Difference approach.”
 - **Descriptive/predictive claim:** An assertion that estimates the value of Y (estimated or predicted) for population P under dimensions X using method M . E.g., “Drawing on a unique Swiss data set (population P) and exploiting systematic anomalies in countries’ portfolio investment positions (method M), I find that around 8% of the global financial wealth of households is held in tax havens (value of Y).”
- **Cleaning code:** A script associated primarily with data cleaning. Most of its content is dedicated to actions like deleting variables or observations, merging data sets, removing outliers, or reshaping the structure of the data (from long to wide, or vice versa).
- **Coding error:** A coding error may occur when a section of the code executes a procedure that contradicts the intended procedure expressed in the documentation (paper or comments of the code). For example, an error happens if the paper specifies that the analysis is performed on the population of males, but the code restricts the analysis to females only.
- **Data availability statement:** A description, normally included in the paper or the appendix, of the terms of use for data used in the paper, as well as the procedure to obtain the data (especially important for restricted-access data). Data availability statements expand on and complement data citations. Find guidance on data availability statements for reproducibility [here.](#)

- **Data citation:** The practice of referencing a dataset, rather than just the paper in which a dataset was used. Data citations help other researchers find data and reward data sharing. Find further guidance on data citations [here](#).
- **Data sharing:** Making the data used in an analysis available to others, ideally through a trusted repository (see below).
- **Declared paper:** The paper that the reproducer analyzes throughout the exercise.
- **Digital Object Identifier (DOI):** A DOI is a string of numbers, letters, and symbols used to permanently identify digital scientific contributions, such as articles, datasets, software, and others (also see DOI.org). Once linked to a contribution, DOIs are permanent and are thus helpful to include in references.
- **Disclosure:** In addition to publicly declaring all potential conflicts of interest, researchers should provide rich details on the methods used for testing a hypothesis, e.g., by including the outcomes of all regression specifications tested. This can be presented in the appendix or supplementary material.
- **Display item:** A display item is a figure or table that presents the results associated with a given research claim found in a research paper.
- **Intermediate data:** Data that has been processed in some fashion but it is not directly used as final input for analyses presented in the final paper (including appendices).
- **Literate programming:** The practice of writing and commenting code in an intertwined fashion, such that both components (code and text) can be read in narrative form.
- **Pre-specification (or pre-registration):** The act of detailing, ahead of time, the statistical analyses that will be conducted for a given research project. Expected outcomes, control variables, and regression specifications are all written in as much detail as possible. This serves to make research confirmatory in nature. Researchers commonly record such details in a hypothesis or trial registry, usually while also providing more details in a pre-analysis plan.
- **Raw data:** Unmodified data files obtained by the authors from the sources cited in the paper. Data from which only personally identifiable information (PII) have been removed are *still considered raw*. All other modifications to raw data make it *intermediate*.
- **(Trial) Registry:** A public database of registered studies or trials, e.g. The American Economic Association’s registry for randomized controlled trials, the Open Science Framework registries, or ClinicalTrials.gov. Some of the largest registries only accept randomized trials, hence the frequent discussion of “trial registries.” *Registration* is the act of publicly declaring that a hypothesis is being, has been, or will be tested, regardless of publication status (see also “Pre-specification” above). Registrations are time-stamped.
- **Replication:** Conducting an existing research project again. A subtle taxonomy exists, and there is disagreement, as explained in Hamermesh, 2007 and Clemens, 2015. *Pure Replication*, *Reproduction*, or *Verification* entails re-running existing code, with error-checking, on the original dataset to obtain the published results. *Scientific Replication* entails attempting to reproduce the published results with a new sample, either with the same code or with slight variations on the original analysis.
- **Reproducibility:** A research paper or a specific display item (an estimate, a table, or a graph) included in a research paper is reproducible if it is possible to reproduce within a reasonable margin of error using the data, code, and materials made available by the author. Computational reproducibility is assessed through the process of **reproduction**.
- **Reproduction package:** A collection of all the materials necessary to reproduce a specific display item or an entire paper. A reproduction package may contain data, code, and other documentation. When the materials are provided in the original publication, they should be labeled as “original reproduction package”; when a reproducer provides them, they should be referred to as “Reproducer X’s reproduction package.”
- **Reproduction tree (or reproduction diagram):** A diagram generated at the Assessment stage on the SSRP, which links display items with the code and data files that are required to reproduce them. The tree is meant to represent the entire computational workflow behind a result from the paper. It can also be used to guide users of the reproduction package and/or to identify missing components for a complete reproduction.

- **Researcher degrees of freedom:** The flexibility a researcher has in data analysis and data cleaning to, consciously or unconsciously, make certain analytical choices to obtain a desired result (most commonly to achieve statistical significance) . This can take a number of forms, including specification searching (also referred to as “p-hacking”), covariate adjustment, selective reporting, or hypothesizing after the results are known (HARKing).
- **Robustness check:** An change in a computational choice, both in data analysis and data cleaning, and its subsequent effect on the main estimates of interest. The SSRP distinguishes between *reasonable specifications* and *feasible specifications*. Reasonable robustness checks (Simonsohn, Simmons, and Nelson 2019) are (i) sensible tests of the research question, (ii) expected to be statistically valid, and (iii) not redundant with other specifications in the set. The set of feasible robustness checks is defined by all the specifications that can be computationally reproduced. We assume that the specifications already published in the paper are part of the set of reasonable specifications.
- **Specification searching:** Searching blindly or repeatedly through data to find statistically significant relationships. While not necessarily inherently wrong, if done without a plan or without adjusting for multiple hypothesis testing, test statistics and results no longer hold their traditional meaning, resulting in false positives.
- **Trusted digital repository:** An online repository whose mission is to provide “reliable, long-term access to managed digital resources to its customers, now and in the future.”(see OCLC (2002)). Storing data here is superior to simply posting on a personal website since it is more easily accessed, less easily altered, and more permanent.

Chang, Andrew, and Phillip Li. 2015. “Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say ‘usually Not’.” *Available at SSRN 2669564*.

Galiani, S, P Gertler, and M Romero. 2018. “How to Make Replication the Norm.” *Nature* 554 (7693): 417–19.

King, Gary. 1995. “Replication, Replication.” *PS: Political Science and Politics* 28: 444–52.

Kingi, Hautahi, Lars Vilhuber, Sylvie Herbert, and Flavio Stanchi. 2018. “The Reproducibility of Economics Research: A Case Study.” In. Presented at the BITSS Annual Meeting 2018; available at the Open Science

National Academies of Sciences. 2019. *Reproducibility and Replicability in Science*. National Academies Press. <https://pubmed.ncbi.nlm.nih.gov/31596559/>.

Simonsohn, Uri, Joseph P Simmons, and Leif D Nelson. 2019. “Specification Curve: Descriptive and Inferential Statistics on All Reasonable Specifications.” *Available at SSRN 2694998*.