

Guidelines for Computational Reproducibility in Economics

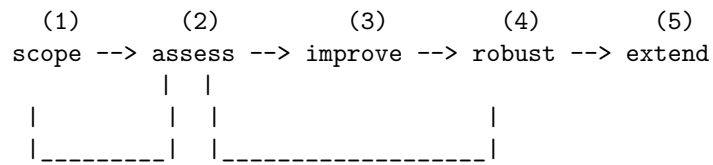
ACRE Team

2020-03-11

Contents

	5
1 Introduction	7
1.1 Beyond binary judgments	8
1.2 Stages of the exercise	8
1.3 Recording the results of the exercise	9
2 Scoping	11
2.1 Steps	11
2.2 Sample strategies for scoping	14
3 Assessment	17
3.1 Describe the inputs.	17
3.2 Connect each output to all its inputs	20
3.3 Assign a reproducibility score.	21
4 Improvements	27
4.1 Types of output-level improvements	27
4.2 Types of paper-level improvements	30
5 Checking for Robustness	33
5.1 Identifying Analytical Choices	33
5.2 Identifying Choice Type	35
5.3 Identifying Analytical Choices	35

5.4	Identifying Analytical Value	35
5.5	Choose and justify alternative values for analytical choices	35
5.6	Test the robustness of results	35
6	Concluding the reproduction	37
6.1	Final products	37
7	Guidance for a Constructive Exchange Between Reproducers and Original Authors	39
7.1	For Reproducers Contacting the Authors of the Original Study .	41
7.2	For Original Authors Responding to Requests from Reproducers	48
8	Code of Conduct	49
9	Contributions	51
9.1	Ask for feedback on guidelines	51
9.2	List of Contributors	51
10	Reproduction Diagrams	53
10.1	Different Scenarios	53
10.2	Additional resources	54
11	Additional resources	57
11.1	Some summaries	57
11.2	Links	57
12	Definitions	59



Scoping

Assessment

Improvement

Robustness

Extensions

Outcome-level

Paper-level

Select paper

Describe inputs

+ Raw data

+ Version control

Analytical choices

New method

Check ACRE

Reproduction diagrams

+ Analysis data

+ Documentation

Type of choice

New data

Search for rep. mat.

Reproduction score

- + Analysis code
- + Dynamic document
- Choice value
- New data
- Read and summ.
- Debug analysis code
- + File structure
- Justify and test alternatives
- Output(s)
- Debug cleaning code
- Depth
- Debug cleaning code
- Record results is Survey 1
- Record results is Survey 2
- Record results is Survey 3

Chapter 1

Introduction

In 2019, the American Economic Association began requiring all the materials for reproduction of any new publication. In addition to the requirements, several specific recommendations were produced to facilitate the authors' compliance. This change in policy and guidance should radically improve the computational reproducibility of all new published research. However all the published research before 2019 did not have neither the requirements nor the guidance to develop proper reproduction materials.

As a result, several studies point towards rates of computational reproducibility in economics that range from alarmingly low (Galiani et al., 2018; Chang and Li, 2015) to just low (Kingi et al., 2018). The project of Advancing Computational Reproducibility in Economics (ACRE) has the goals of improving, and having a much clearer picture of, computational reproducibility in economics.

Replication, or the process by which a study's hypotheses and findings are re-examined using different data or different methods (or both) (King 1995) is an essential part of the scientific process that allows science to be "self-correcting." *Computational reproducibility*, or the ability to reproduce the results, tables, and other figures using the available data, code, and materials, is a precondition for replication. Computational reproducibility is assessed through the process of *reproduction*. At the center of this process is the *reproducer* (you!), a party not 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.

This exercise is designed for reproductions performed in Economics graduate courses or undergraduate theses, with the goal of providing common terminology and standards for conducting reproductions. The goal of reproduction, in general, is to assess and improve the computational reproducibility of published research in a way that facilitates future replication, extension, and collaboration.

This exercise is part of the Advancing Computational Reproducibility in Economics (ACRE) project led by the Berkeley Initiative for Transparency in the Social Sciences (BITSS) and Prof. Lars Vilhuber, Data Editor for the journals of the American Economic Association (AEA). The ACRE project builds on efforts to improve the reproducibility of economics research through the updated Data and Code Availability Policy, which since 2019 mandates pre-publication verification of computational reproducibility by the AEA Data Editor. Several specific recommendations were also produced to facilitate the authors' compliance.

1.1 Beyond binary judgments

Assessments of reproducibility can easily gravitate towards binary assessments that declare a entire paper “reproducible” or “non-reproducible”. These guidelines suggest a more nuance approach by highlighting three dimensions that make binary judgement less relevant. First one paper usually contains several scientific claims, and those claims can vary in computational reproducibility. Second, each claim is typically measured through several specifications. This makes easier for reproducers to hypothesize after the fact about which are the main specifications where reproducibility assessed. Finally, there are several levels of reproducibility, ranging from the absence of any materials to the complete reproducibility starting from the raw data. And even for a specific claim-specification, distinguishing the appropriate level can be far more constructive than simple labeling as (ir)reproducible.

It should be noticed that the highest level of reproducibility, which requires complete reproducibility starting from the raw data, is very demanding and it should not be expected of published research (especially that published before 2019). This high bar is set up as an aspiration to improve the current reproducibility of research and facilitate the transmission of knowledge in the scientific community.

1.2 Stages of the exercise

This reproduction exercise is divided into four stages, with a fifth optional stage :

1. **Scoping**, where the reproducer defines the scope of the exercise by declaring a paper and specific output(s) on which they will focus;
2. **Assessment**, where the reproducer reviews and describes in detail the available reproduction materials (or the “reproduction package”) and

paper of choice and specific outputs you will attempt to reproduce. This step also involves writing a brief 1-2 page summary of your chosen paper. In assessment, you will inspect the paper's reproduction materials (raw data, analysis data, and code), draw diagrams connecting the **output to be reproduce** with its inputs, and score the level of reproducibility. All this information will be recorded in a standardized spreadsheet. In improvement, a second short survey will ask you to record specific improvements and report potential increases in the level of reproducibility. Finally, in robustness checks you will record any analytical choices and their possible variations.

Chapter 2

Scoping

In the Scoping stage, you will choose a paper to reproduce. , or be assigned one. Chapter 2 describes the steps for defining the scope of a reproduction exercise. *Use Survey 1 to record your work for this stage.*

2.1 Steps

Before you invest any time in reading or summarizing a paper, complete steps 2.1.1–2.1.2:

2.1.1 Check records on the ACRE platform.

Check the ACRE database for previous assessments of the same paper. If there are none, create the first entry. If there are previous entries, check if the **paper status** is “open to reproductions”, and create a new entry¹.

2.1.2 Verify that reproduction materials exist.

Verify that the paper has reproduction materials (consult the unofficial AEA Data and Code Guidance to determine whether you have everything you need before you start). If any of the materials are missing, use our records to check if requests to authors for these materials have been made before. If nobody has yet submitted a similar request, contact the authors using language suggested here, making sure to copy (acre@berkeley.edu). If the authors have been contacted

¹If the paper status is “closed to reproductions”, ACRE has documented interactions between previous reproducers and original authors, so it is not possible to access any further required materials.

before and the information is still missing, submit a request for closing the paper for further reproductions and choose another paper.

2 week

1 month

1 semester

analysis data

raw data

analysis data

raw data

analysis data

raw data

scoping

10% (1 day)

5% (1 day)

3% (3 days)

Assessment

25%

15%

7%

Improvement

20%

0%

40%

20%

30%

Robustness

40%

30%

30%

Extension

5%

0%

10%

10%

2.1.3 Read and summarize the paper.

Depending on how much time you have, we recommend that you write a short summary of the paper. This will help remind you of the key elements to focus on for the reproduction, and to demonstrate your understanding of the paper (for yourself and potentially for the original author).

Read the paper and write a two-page summary, answering the following: What is the main question the paper attempts to answer? What methodologies does it employ? What data sources do the authors use?

2.1.4 Declare scope of the exercise.

Declare if your assessment will be of all of the paper's results (**complete**) or just the **main results**, and if you will aim to start with **raw data** or **processed data**. Find definitions of these concepts below.

ADD description that scoping involves choosing paper, output and level of depth to improve. Given output assessment is all the way to raw data. But improvement and robustness should be limited to the amount of time is provided for the exercise.

2.1.4.1 Define your reproduction margins.

2.1.4.1.1 Define the *extensive margin*: outputs to reproduced

Define whether you will attempt to reproduce all of your chosen paper's results or just the main results. - **Complete**: A successful complete reproduction obtains all of the same estimates presented in the chosen paper. This includes tables, figures, and inline estimates, both in the main body of the paper and the appendices. - **Main results**: A successful reproduction of main results obtains the same estimates as those highlighted in the chosen paper's abstract, introduction, or conclusion. If no estimate is highlighted, you should choose the main result and declare it at the beginning of the reproduction.

2.1.4.1.1.1 Describe the types of output.

Broadly describe the outputs you will attempt to reproduce.

- **Tables**: Describe how to cite record location.

– **Figures:** ADD

– **In-line:** ADD

2.1.4.1.2 Define the *intensive margin*: depth of reproduction in terms of original data.

- **Raw data** sets are unmodified files obtained by the authors from the sources cited in the paper. Raw data from which personally identifiable information (PII) has been removed is still considered raw. All other modifications to raw data make it *processed*.
- **Processed data** are raw data sets that have gone through any transformation other than the removal of PII. There are two kinds of processed data:
 - *Intermediate data* are not directly used as final input for analyses presented in the final paper (including appendices). Intermediate data should not contain direct identifiers.
 - *Analytic data* is used as the final input in a workflow in order to produce a statistic displayed in the paper (including appendices).

2.1.4.1.2.1 Classifying a data set

A data set may be classified as **raw** if it fits any of the following criteria:

- The data is stored in a folder or file labeled as “raw”.
- The data set is not the output of any code script in the reproduction materials.
- The same data file can be independently obtained from the data source cited in the paper.

A data set may be classified as **analytic** if it fits any of the following criteria:

- The data is stored in a folder or file labeled as “analytic” or “analysis.”
- The data set is the last input required to produce some of the output (formatted or unformatted) of the paper.

2.2 Sample strategies for scoping

ADD.

2.2.1 Strategy 1:

1. Attempt to reproduce the paper's main estimates.
2. Attempt to reproduce two main tables and/or two main figures.
3. Try to make improvements to these outputs.
4. Try to reproduce the rest of the paper.

2.2.2 Strategy 2:

1. Attempt to reproduce all outputs resulting from the analytic data.
2. Once you can do that, try to make improvements.
3. If you can make improvements, attempt to reproduce the outputs starting from raw data. (or in parallel?)

Chapter 3

Assessment

The goal of this stage is to conduct a standardized assessment of your chosen paper’s computational reproducibility. This stage is designed to record as much of the learning process behind a reproduction as possible. Such a record facilitates future incremental improvements, allowing new reproducers to pick up where others have left off.

First, you will provide a detailed description of the reproduction materials. Second, you will connect the outputs you’ve chosen to reproduce with their corresponding inputs. With these elements in place, you can score the level of reproducibility of each output, and report on paper-level dimensions of reproducibility.

Use Survey 2 to record your work as part of this step.

Tip: We recommend that you first focus on one specific output (e.g. “Table 1”). After completing the assessment for this output, you will have a much easier time translating improvements to other outputs.

ADD: how a reproduction package should look like

3.1 Describe the inputs.

This section explains how to list all the input materials found or referred to in the reproduction package. First, you’ll identify data sources and connect them with their raw data files (when available). Second, you’ll locate and provide a description a brief description of the analytic data files. Finally, you’ll locate, inspect, and describe the analytic code used in the paper.

The following terms will be used in this section:

- **Cleaning code:** A script associated primarily with data cleaning. Most of its content is dedicated to actions such as deleting variables or observations, merging data sets, removing outliers, or reshaping the structure of the data (from long to wide or vice versa).
 - **Analysis code:** A script associated primarily with analysis. Most of its content is dedicated to actions such as running regressions, running hypothesis tests, computing standard errors, and imputing missing values.
 - ADD DEFINITION FOR REPRODUCTION PACKAGE.

3.1.1 Describe the data sources and raw data.

In the paper to be reproduced, find references to all data sources used in the analysis. A data source is usually described in narrative form. For example, if the body of the paper uses text like “...for earnings in 2018 we use the Current Population Survey...”, the data source is “Current Population Survey 2018”. If it is mentioned for the first time on page 1 of the Appendix, its location should be recorded as “A1”. Do this for all data sources mentioned in the paper.

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), while others are less clear (e.g., “our information on regional minimum wages comes from the Bureau of Labor Statistics.” This should be recorded 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 available materials. Record their folder locations relative to the main reproduction folder. In addition to looking at the existing data files, we recommend you review the first lines of all code files (especially cleaning code), looking for lines that call the data sets. Inspecting these scripts may help you understand how different data sources are used, and possibly identify any files missing from the reproduction package.

Record this information in a standardized spreadsheet with the following structure:

Raw data information:

data_source	page	data_files
"Current Population Survey 2018"	A1	cepr_march_2018.dta
"DHS 2010 - 2013"	4	nicaraguaDHS_2010.csv; boliviaDHS_2010.csv; nicaraguaDHS_2011.csv;

		nicaraguaDHS_2012.csv; boliviaDHS_2012.csv;	
		nicaraguaDHS_2013.csv; boliviaDHS_2013.csv	
"2017 SAT scores"	4	Not available	
...

3.1.2 Describe the analytic data sets.

List all the analytic files you can find in the reproduction materials, and identify their locations relative to the main reproduction folder¹. Record this information in a standardized spreadsheet.

As you progress through the exercise, add to the spreadsheet a one-line description of each file's main content (for example: `all_waves.csv` has the simple description `data for region-level analysis`). This may be difficult in an initial review, but will become easier as you go along.

The resulting report will have the following structure:

Analysis data information:

analysis_data	location	description
final_data.csv	/analysis/fig1/	data for figure1
all_waves.csv	/final_data/v1_april/	data for region-level analysis
...

3.1.3 Describe the code scripts.

List all code files that you find in the reproduction materials and identify their locations relative to the master reproduction folder. Review the beginning and end of each code file and identify the inputs required to successfully run the file. Inputs may include data sets or other code scripts that are typically found at the beginning of the script (e.g., `load`, `read`, `source`, `run`, `do`). Outputs may include other data sets or plain text files that are typically at the end of a script (e.g., `save`, `write`, `export`). Record this information in the standardized spreadsheet.

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

As you gain an understanding of each code script, you will likely find more inputs and outputs – we encourage you to update the standardized spreadsheet. Once finished with the reproduction exercise, classify each code file as *analysis* or *cleaning*. This subjective assessment should be made based on your interpretation of each script’s main role.

Record this information in the standardized spreadsheet

List any input files (data and code) required to execute this file successfully. Separate each item with “;”.

List all outputs generated with this file. This includes data sets, figures and tables, and objects used later on. Separate each item with “;”.

Provide a one-line description of what each script does.

Classify the code as “primarily cleaning code” or “primarily analysis code”.

file_name	location	inputs	outputs
output_table1.do	/code/analysis/	analysis_data01.csv	output1_part1.txt
data_cleaning02.R	/code/cleaninig/	admin_01raw.csv	analysis_data02.c
...

3.2 Connect each output to all its inputs

Draw diagrams connecting outputs to raw data sources. To do this, find the code script that generates the target output (formatted or not), then find all inputs required to execute the script (including data and other code files). Repeat until you reach the raw data or the last available file.

When a connection cannot be drawn due to a missing component, use the “-||-” symbol. For more examples of diagrams connecting final output to initial raw data, see [here](#).

```

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

```

This diagram can be represented in data format by specifying how each component depends on its inputs. For example:

Data representation of diagram behind Table 1.

ouput	order	component	depends_on	inpt_type
table1	1	table1	formatting_table1.R	code
table1	2	formatting_table1.R	output1_part2.txt	output
table1	3	formatting_table1.R	output1_part1.txt	output
table1	4	output_table1.do	analysis_data01.csv	data
...

Record this information in the standardized spreadsheet. If you have difficulty translating the diagram into a spreadsheet, you can draw it with pen and paper, take a picture and upload it to the assessment survey.

3.3 Assign a reproducibility score.

Once all possible inputs have been identified, and there is a clear understanding of the connection between the outputs and inputs, you can start to assess the output-specific level of reproducibility.

The following concepts will be used in this section:

- **Computationally Reproducible from Analytic data (CRA):** The output can be reproduced with minimal effort starting from the analytic data sets.
- **Computationally Reproducible from Raw data (CRR):** The output can be reproduced with minimal effort from the raw data sets.

- **Minimal effort:** Five minutes or less are required to run the code, not including computing time.

3.3.1 Levels of Computational Reproducibility for a Specific Output

Here we outline different levels of computational reproducibility. Each level is defined by data and code availability, possible improvements, and whether or not they achieve some type of reproducibility. In the next chapter we'll describe each possible improvement in more detail.

MAKE EXPLICIT UNDERLYING VALUATIONS:

- L1: The worst is to have no access to any data or code.
- L2: Having only code is better than nothing, but worst than any other combination
- L3: Having analytic data but not code is better than having code and no data.
- L4 and L5: Conditional on having all analytic material. Reproducible is better than not
- L6: Having raw and analytic data is better than having code but no data.
- L7: Given that there is raw and analysis data, having cleaning code only does not add much. having analysis code does add (to lvl 7 and lvl 8 if CRA)
- L9: Having all materials that but not being able to reproduce is better than all previous cases. Achieving CRA (lvl 10) and CRR (lvl 11) are the best levels according to this criteria.

ADD ONE QUESTION AT THE END OF SURVEY TWO TO ASK REPRODUCER IF THEY AGREE WITH THIS ORDER, AND IF THEY WOULD ADD/DELETE/MODIFY ANY SPECIFIC LEVEL

The assessment is made at the output level – a paper can be highly reproducible for its main results, but suffer from low reproducibility for other outputs. The assessment is made using a 10-point scale, where 0 represents that, under current circumstances, reproducers cannot access any reproduction materials, while 10 represents access to all the materials and that the target outcome can be reproduced starting from raw data.

Level 1 (L1): No data or code are available. Possible improvements include adding: raw data (+AD), analysis data (+RD), cleaning code (+CC), and analysis code (+AC).

Level 1.5 (L1.5): Code scripts are available, but no data are available. Possible improvements include adding: raw data (+AD) and analysis data (+RD).

Level 2 (L2): Analytic data are available, but raw data and code are not. Possible improvements include adding: raw data (+RD) and analysis code (+AC).

Level 3 (L3): Analytic data sets and analysis code are available. However, code does not run or produces different results than those in the paper (not CRA). Possible improvements include obtaining raw data (+RD) or debugging the analysis code (DAC).

Level 4 (L4): Analytic data sets and analysis code are available. They produce the same results as presented in the paper (CRA). The reproducibility package may be improved by obtaining the original raw data sets, or by documenting the steps required to obtain those files.

Level 5 (L5): Analytic data and raw data are available. However, some or all cleaning and analysis code are missing. Possible improvements include: adding analysis code (+AC) and/or cleaning code (+CC).

Level 6 (L6): All data and analysis code are available. However, the code does not run or produces different results than those presented in the paper (not CRA). Possible improvements include: adding the missing cleaning code (+CD) or debugging the analysis code (DAC).

Level 7 (L7): All data and analysis code are available, but cleaning code may be missing. They produce the same results as presented in the paper (CRA). The reproducibility package may be improved by adding the missing cleaning code (+CD).

Level 8 (L8): All materials (raw data, analytic data, cleaning code, and analysis code) are available. However, the code does not run or produces different results than those presented in the paper (not CRR and not CRA). Possible improvements include: debugging the cleaning code (DCC) or debugging the analysis code (DAC).

Level 9 (L9): All 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 does not run or produces different results than those presented in the paper (not CRR). Possible improvements include: debugging the cleaning code (DCC).

Level 10 (L10): All materials are available and produce the same results as presented in the paper with minimal effort, starting from the analytic data (yes CRA) or the raw data (yes CRR).

The following figure summarizes the different levels of computational reproducibility (for any given output). For each level, there will be improvements that: have been made (-), can be made to move up one level of reproducibility (), or are out of reach given the current level of reproducibility (x).

Figure 1: Levels of Computational Reproducibility and Possible Improvements

		Possible improvements				
		Level	+Analysis Data	+Raw Data	+Analysis Code(AC)	+Clean Code
What data are available?						
None	L1					
Analytic data only. Code?						
No code or cleaning code only.....	L2	-				
Analysis code only. Is it CRA?						
No.....	L3	-			-	
Yes.....	L4	-			-	
Raw & Analytic data. Code?						
None	L5	-	-			
Analysis code only. CRA?						
No.....	L6	-	-		-	
Yes.....	L7	-	-		-	
A. and cleaning code. Is it CRR?						
No. CRA?						
No.....	L8	-	-		-	-
Yes.....	L9	-	-		-	-
Yes.....	L10	-	-		-	-

Choose the appropriate level of computational reproducibility and record it using the following format.

output_name	level	additional_explanation	other_info
table 1	4		...
table 2	7		...
figure 1	5		...
...

Record this information in the standardized spreadsheet. You will be asked to provide this information in the assessment and improvement survey.

3.3.2 Reproducibility dimensions at the paper level

In addition to an output-specific assessment of computational reproducibility, several practices can facilitate a paper's overall computational reproducibility.

These practices are described in detail in the Improvement chapter. In the Assessment section, it is only required that you verify whether the original reproduction package made use of any of the following:

- master script that runs all steps
- readme file
- standard file organization
- version control
- analysis uses open source software
- dynamic document
- computing capsule (e.g. CodeOcean, Binder, etc.)

Congratulations! You have now completed the Assessment stage. You have provided a concrete building block of knowledge to improve understanding of the state of reproducibility in Economics.

Please continue to the next section where you can help to improve it!

Chapter 4

Improvements

After completing an assessment of the paper’s reproducibility packages, you can start proposing ways to improve it’s reproducibility. Making improvements provides an opportunity to gain deeper understanding of the paper’s methods, findings, and overall contribution. Each contribution can also be assessed and used by the wider ACRE community—other students and researchers using the ACRE platform around the world.

As in the Assessment section, we recommend that you first focus on one specific output (e.g., “Table 1”). After making improvements to this first output, you will have a much easier time translating those improvements to other outputs.

Use Survey 2 to record your work as part of this step.

4.1 Types of output-level improvements

4.1.1 Adding missing raw data files or meta-data (+RD)

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 a specific missing file. During Assessment, you identified all data sources from the paper’s body and appendices (column `data_source` in this standardized spreadsheet). However, some data sources (as collected by the original investigators) might be missing one or more data files. You can sometimes find the specific name of those files by looking at the beginning of the cleaning code scripts. If you find the name of the file, record it in the `data_file` field of the same spreadsheet as above. If not, recorded it as “Some (or all) of the files used in the paper corresponding

to data source X”.

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 the reproduction package. Make sure to obtain permission from the original author to repost this data. See tips for communication for a template email.
 - 2.2 - If no: proceed to step 3.
3. Use the ACRE database to verify whether there have been previous attempts to contact the authors regarding this paper.
4. Contact the original authors and politely request the original materials. Be mindful of the authors’ 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 tips for communication for sample language on how to approach authors.
5. If the data sets are not available due to confidentiality or proprietary issues, you can still improve the reproduction package by providing detailed instructions for future researchers to follow, including contact information and possible costs.

In addition to trying to obtain raw data, you can also contribute by obtaining missing analytic data.

4.1.2 Adding missing analytic data files (+AD)

Analytic data can be missing for two reasons: (i) raw data exists, but the procedures to transform it into analytic data are not fully reproducible, or (ii) 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 (eg `analysis_data_03.csv`).
2. Verify that the data cannot be obtained by running the data cleaning code over the raw data.
3. Use the ACRE database to verify if previous attempts have been made to contact the authors about this data.
4. Contact the authors and request the specific data set.

4.1.3 Adding missing analysis code (+AC)

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 no relevant code file is found, identify the location of the missing file relative to the steps in the reproduction diagram.
3. Use the ACRE database to verify if previous attempts have been made to contact the authors about this issue.
4. Contact the authors and request the specific code files.
5. If the authors don’t respond, 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 (+CC)

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 (1-5).

4.1.5 Debugging analysis code (DAC)

Whenever code is available in the reproduction package, you should be able to debug those scripts. There are four 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 instructions with new ones that perform the same tasks but take less time (e.g., choose one numerical optimization algorithm over another, but obtain the same results).

- *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 happen if the paper specifies that the analysis is performed on a population of males, but the code restricts the analysis to females only. Please follow the ACRE procedure to report coding errors.

4.1.6 Debugging cleaning code (DCC)

Use the same steps as you took to debug the analysis code, but report them separately.

4.1.7 Reporting results

Track all the different types of improvements you make and record in this standardized spreadsheet with the following structure:

Level-specific quality improvements: add data/code, debug code.

output_name	imprv	description_of_added_files	lvl
table 1	+AD	ADD EXAMPLES	5
table 1	+RD	ADD EXAMPLES	5
table 1	DCC	ADD EXAMPLES	5
figure 1	+CC		6
figure 1	DAC		6
inline 1	DAC		8
...

4.2 Types of paper-level improvements

There are at least six additional improvements you can make to improve a paper's overall reproducibility. These additional improvements can be applied across all reproducibility levels (including level 10).

1. Set up the reproduction package using version control software, such as Git.

2. Improve documentation by adding extensive comments to the code.
3. Integrate the documentation with code by adapting the paper into a literate programming environment (e.g., using Jupyter notebooks, RMarkdown, or a Stata Dynamic Doc).
4. If the code was written using a proprietary statistical software (e.g., Stata or Matlab), re-write it using an open source statistical software (e.g., R, Python, or Julia).
5. Re-organize the reproduction materials 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.
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.

4.2.1 Reporting improvements

You will be asked to provide this information in the Assessment and Improvement Survey.

Chapter 5

Checking for Robustness

[UNDER CONSTRUCTION]

- Identify all possible analytical choices: original and repeated ones.
- Identify type of choice.
- Identify choice value.
- Suggest choice alternative and justify (one line)

5.1 Identifying Analytical Choices

As part of the requirements to demonstrate comprehension of the paper and the code researchers conducting the reproduction will be asked to record all the analytical choices identified during the code review process. This is done in two steps: first adding comment lines into the code files where an analytic choice are found, and second, compiling those analytic choices into a standardized data set.

In your copy of the replication code, add the comment `"# ANALYTICAL CHOICE OF TYPE ____ . RECORDED FOR THE FIRST TIME [HERE or IN "FILE_NAME-LINE_NUMBER"]"` above each analytical choice detected in the code. Possible types of analytical choices include (but are not limited to):

- Analytical choices in data cleaning code:
 - Variable definition
 - Data sub-setting

- Data reshaping (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, etc.)
 - Controls
 - Adjustment of standard errors
 - Choice of weights
 - Treatment of missing values
 - Imputations
 - Other (specify as “methods - other”)

Once finished, transcribe all the information on analytical choices into a data set. For the `source` field type “original” whenever the analytical choice is identified for the first time, and `file_name-line number` every time that the same analytical choice is applied subsequently (for example if a analytic choice is identified for the first time in line 103 and for a second in line 122 their respective values for the `source` field should be `original` and `code_01.do-L103` respectively).

The resulting data base should have the following structure:

file_name	line_number	choice_type	choice_value	Source
code_01.do	73	data	males	original
		subsetting		
code_01.do	122	variable	income = wages +	“code_01.do-
		definition	capital gains	L103”
code_05.R	143	controls	age, income,	original
			education	
...

5.2 Identifying Choice Type

5.3 Identifying Analytical Choices

5.4 Identifying Analytical Value

5.5 Choose and justify alternative values for analytical choices

5.6 Test the robustness of results

Test the robustness of results to alternative (sensible) specifications

- Identify sensible alternatives to analytical choices.
 - Sample from sensible analytical choices and re-run: report how much do results change as fraction of standard deviations.
 - Jackknife the preferred estimate.
 - Use ML to select among covariates...

Chapter 6

Concluding the reproduction

[UNDER CONSTRUCTION] but testing now

Walk the students on checking that they have completed all the steps and where can they see their output.

6.1 Final products

- One-page introduction describing why you chose this paper
- Two-page summary of paper
- 2 Completed surveys:
 - i - General information about the paper and specific information about output to reproduce.
 - ii - Assessment of how (computationally) reproducible is the paper; description of improvements to its reproducibility; record of all the analytical choices identified in the exercise.
- ACRE report card with all the improvements that were created by the researcher reproducing the paper. The list of improvements will be made public and original authors will receive a copy of the report card. The option of anonymity will be provided to the researchers reproducing the paper.
- New Readme file (autogenerated).

- Data with all analytical choices identified.
- ?? Narrated description of improvements to original CR of the paper, assessment of robustness of results. Lessons from the exercise and possible future extensions.

Chapter 7

Guidance for a Constructive Exchange Between Reproducers and Original Authors

The purpose of this chapter is to facilitate constructive and respectful communication between reproducers (you) 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 their interactions with original authors are constructive. We provide a summary 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 the original work.
- b. Explain why you have chosen this study. Try to proactively address potential concerns for selection bias.
- c. 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 would have done for your own work.

- a. Avoid binary judgments and statements like “failed to reproduce.” Clearly state 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 put these recommendations into practice, we provide template language for common scenarios that reproducers and authors may encounter in their interactions.

For reproducers:

- A. Contacting the original author(s) to request items that are missing;
- B. Asking for additional guidance when some materials have been shared;
- C. Response to an original author who has refused to share an item due to undisclosed reasons;
- D. Response to an original author who has refused to share data due to legal or ethical constraints;
- E. Contacting the original author(s) to share the results of your reproduction;
- F. Responding to adversarial responses from original authors.

For authors:

- G. Responding to a request [TODO]
- H. Acknowledging that some information is missing [TODO]
- I. Acknowledging that some material is still embargoed for future research [TODO]
- J. Responding to incomplete/aggressive requests from reproducers [TODO]

While we hope that you find these useful, note that they are *only recommendations*, and you are welcome to modify them based on the context and needs

of your specific project. Feel free to contact us if you need more guidance or would like to provide feedback on these materials.

7.1 For Reproducers Contacting the Authors of the Original Study

7.1.1 Before you contact the original author

Considering the following:

1. 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. All papers published in AEA journals after July 2019 should have a Readme file.
 - Check whether there are any restrictions to accessing the data or code, and whether there are instructions on how to access these files for the purpose of reproduction.
2. If reproduction materials are not readily available in the location where the article is published (e.g., the journal website), check the authors' websites, Dataverse profiles, the ICPSR Publications Related Archive, and other relevant archives and/or data repositories.
3. If steps 1 and 2 don't work, 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 being reproduced (include version, date, and a DOI link (or just a URL));
 - Context for the reproduction (as part of a class exercise, thesis, etc.) and a notice that the outcome will be recorded in the ACRE reproducibility database;
 - Items from the reproduction package that are missing, as well as locations where you had (unsuccessfully) searched for them;
 - 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 exercise with the original authors?
 - A deadline to respond (we suggest at least two weeks).
4. 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.

7.1.2 A. Contacting the original author(s) to request items that are missing

Template email:

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

Dear Dr. [Lastname of Corresponding Author],

I am contacting you about reproduction materials for your paper titled [Title] which 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 to reproduce this paper as part of a class exercise. [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 be able to 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 (raw/analytic data, code, protocols for conducting the experiment, etc.)]. I have already searched [locations where you searched for items, with links provided], however I was not able to locate the items there. You can be assured that I will not share any of the materials without your permission, and I will use them exclusively for the purpose of this exercise. Let me know if there are any legal or ethical restrictions that apply to all or parts of the reproduction materials so that I can take that into consideration during this exercise.

Note that I will record the outcome of my reproduction on the Advancing Computational Reproducibility in Economics (ACRE) platform [ADD LINK], an online catalog of reproduction projects in economics. ACRE is hosted by the Berkeley Initiative for Transparency in the Social Sciences (BITSS). Let me know if you would

7.1. FOR REPRODUCERS CONTACTING THE AUTHORS OF THE ORIGINAL STUDY⁴³

like me to share the outcome of my exercise 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]

7.1.3 B. Asking for additional guidance when some materials have been shared

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

Template email:

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

Dear Dr. [Lastname of Corresponding Author],

Thank you for sharing the materials. They have been immensely helpful for my work.

Unfortunately, I ran into a few issues as I delved into the reproduction exercise, and I think your guidance would be helpful in resolving 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 B.1 and B.2 below for more details].

Thank you in advance for your help.

Best regards,
[Reproducer]

7.1.3.1 B.1 An example of well described issues:

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

OUTPUT X

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

I have marked with an asterix (*) 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 be able to identify the name of the raw data set required to obtain output1_part1.txt. This is to let you know that and 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 or somebody in your research group, but I want to assure you that I will add these missing files to the reproducibility package for your paper on the ACRE platform. **Doing this will ensure that you will not be asked twice for the same missing file.**

7.1.3.2 B.2 An example of poorly described issues:

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 a basic principle of open science, so I am confident that you will do the right thing.

7.1.4 C. Response when the original author has refused 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 Dr. [Lastname 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 ACRE platform. I will also post my assessment of the reproducibility of the paper in its current form based on the ACRE reproducibility scale [ADD LINK].

Let me know if you have any questions.

Best regards,

<>

7.1.5 D. Response when the original author has refused to share due to legal or ethical restrictions of the data

Template email:

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

Dear Dr. [Lastname of Corresponding Author],

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

Though I understand that 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 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 (find an example/checklist here[ADD EXAMPLE]); 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 ACRE reproducibility scale [ADD LINK], 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. Let me know if that would be helpful.

Thank you for your help!

Best regards,
[Reproducer]

7.1.6 E. Contacting the original author to share the results of your reproduction exercise

[WORK IN PROGRESS]

7.1.7 F. 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.

7.1.7.1 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 exercise.

Template email:

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

Dear Dr. [Lastname of Corresponding Author],

Thank you for your response. I will work to reproduce using the available materials, and will record my results accordingly on the ACRE platform. I will also post my assessment of the reproducibility

7.1. FOR REPRODUCERS CONTACTING THE AUTHORS OF THE ORIGINAL STUDY⁴⁷

of the paper in its current form based on the ACRE reproducibility scale[[ADD LINK](#)].

Let me know if you have any questions.

Best regards,
[Reproducer]

7.1.7.2 Harassment and/or discrimination

The AEA and other economic 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 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 for the purpose of filing 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 on the basis of 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 are 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.** 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 exercise would be appropriate.

draft here

7.2 For Original Authors Responding to Requests from Reproducers

7.2.1 Responding to a repeated request

[TO DO] ### Acknowledging that some information is missing

[TO DO]

7.2.2 Acknowledging that some material is still embargoed for future research

[TO DO]

7.2.3 Responding to incomplete/aggressive requests from reproducer

Chapter 8

Code of Conduct

[UNDER CONSTRUCTION]

Chapter 9

Contributions

TO DO

9.1 Ask for feedback on guidelines

9.2 List of Contributors

Chapter 10

Reproduction Diagrams

10.1 Different Scenarios

JOEL: Please fill-in.

10.1.1 Complete

```
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
```

10.1.2 Raw data and analytic data, but cleaning code is missing.

```
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] MISSIN FILE(S)
|           [data] admin_01raw.csv

```

10.2 Additional resources

Create a section with short summaries of great resources for comp. repro and invite reader to contribute.

10.2.1 Some summaries

10.2.2 Summary on reproducible workflow (Chapter 11) from Christensen et al. (2019):

- TODO

10.2.3 Links

- Project TIER
- IDB's cheatsheet for transparency, reproducibility and ethics
- Lars Vilhuber LDI's Wiki for Reproducibility. Particularly this section.
- World Bank DIME's Wiki for transparent and reproducible research.
- Dynamic documents in R, Python and Stata
- Git resources:
 - Jenny Bryan's book and video
 - Github learning lab
 - Udacity's intro
 - Git for poets
 - Combining GitHub and Dropbox
 - Atlassian intro to Git

- Software Carpentry tutorial from the command line
- Open Science Framework (OSF)
- R for Stata users

Chapter 11

Additional resources

- **Coding errors:** A coding error will occur when a section of the code, of the reproduction package, executes a procedure that is in direct contradiction with the intended procedure expressed in the documentation (paper or comments of the code). For example an error happens if the paper specify that the analysis is perform on the population of males, but the code restricts the analysis to females only. Please follow the ACRE procedure to report coding errors.

Create a section with short summaries of great resources for comp. repro and invite reader to contribute.

11.1 Some summaries

11.1.1 Summary on reproducible workflow (Chapter 11) from Christensen et al. (2019):

- TODO

11.2 Links

TODO: Add and classify

- Project TIER
- IDB's cheatsheet for transparency, reproducibility and ethics

- Lars Vilhuber LDI's Wiki for Reproducibility. Particularly this section.
- World Bank DIME's Wiki for transparent and reproducible research.
- Dynamic documents in R, Python and Stata
- Git resources:
 - Jenny Bryan's book and video
 - Github learning lab
 - Udacity's intro
 - Git for poets
 - Combining GitHub and Dropbox
 - Atlassian intro to Git
 - Software Carpentry tutorial from the command line
- Open Science Framework (OSF)
- R for Stata users

Chapter 12

Definitions

TO DO

Bibliography

Chang, A. and Li, P. (2015). Is economics research replicable? sixty published papers from thirteen journals say 'usually not'. *Available at SSRN 2669564*.

Christensen, G., Freese, J., and Miguel, E. (2019). *Transparent and reproducible social science research: How to do open science*. University of California Press.

Galiani, S., Gertler, P., and Romero, M. (2018). How to make replication the norm. *Nature*, 554(7693):417–419.

Kingi, H., Vilhuber, L., Herbert, S., and Stanchi, F. (2018). The reproducibility of economics research: A case study. Presented at the BITSS Annual Meeting 2018 and available at the Open Science