BERKELEY INITIATIVE FOR TRANSPARENCY
IN THE SOCIAL SCIENCES

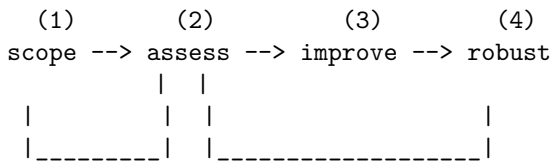
Guidelines for Verification of Computational Reproducibility in Economics

ACRE Team

2020-06-26

Contents

See a full list of contributors



(1) Scoping

(2) Assessment

(3) Improvement

(4) Robustness

(5) 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

Check Rep. pkg exists

Introduction

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. In addition to the requirements laid out in the policy, several specific recommendations were produced to facilitate compliance. This change in policy is expected to improve the computational reproducibility of all published research going forward, after several studies showed that rates of *computational reproducibility* in economics at large range from somewhat low to alarmingly low (**???**).

Replication, or the process by which a study's hypotheses and findings are reexamined using different data or different methods (or both) (**?**) 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 of a paper 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 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.

This exercise is designed for reproductions performed in economics graduate courses or undergraduate theses, with the goal of providing a common approach, 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 further robustness checks, extensions, collaborations, and replication.

This exercise is part of the Accelerating Computational Reproducibility in Economics (ACRE) project, which aims to assess, enable, and improve the computational reproducibility of published economics 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 the Laura and John Arnold Foundation.

Beyond binary judgments

Assessments of reproducibility can easily gravitate towards binary judgements that declare an entire paper "reproducible" or "non-reproducible." These guidelines suggest a more nuanced approach by highlighting two realities 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 **??** illustrates this idea.

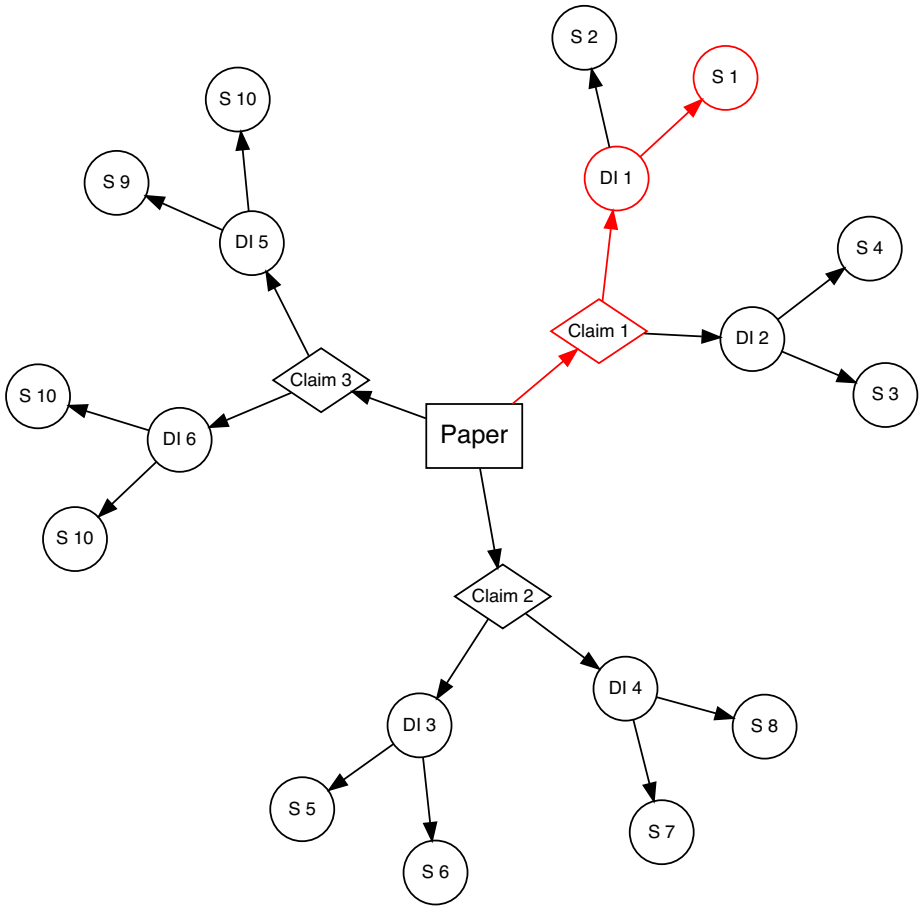


Figure 1: One paper has multiple components to reproduce. DI: Display Item, S: Specification

Second, for any given specification there are several levels of reproducibility, ranging from the absence of any materials to complete reproducibility starting from raw data. And even for a specific claim-specification, 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 to achieve and should not be expected of all published research — especially before 2019. Instead, this level can serve as an aspiration for the field of economics at large as it seeks to improve the reproducibility of research and facilitate the transmission of knowledge throughout the scientific community.

Stages of the exercise

This reproduction exercise is divided into four stages, corresponding to the first four chapters of these guidelines, with a fifth optional stage:

1. **Scoping**, where you (the reproducer) will define the scope of the exercise by declaring a paper and the specific output(s) on which you will focus for the remainder of the exercise;

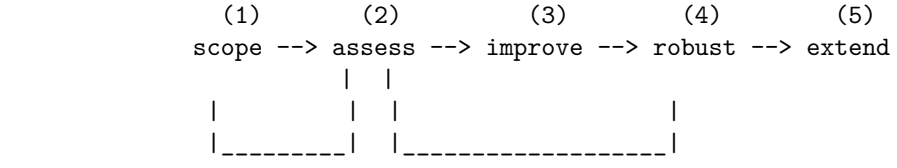
2. **Assessment**, where you will review and describe in detail the available reproduction package, and assess the current level of computational reproducibility of the selected outputs;

3. **Improvement**, where you will modify the content and/or the organization of the reproduction package to improve its reproducibility;

4. **Robustness checks**, where you will assess the quality of selected analytical choices; and

5. **Extension** (if applicable), where you may extend the current paper by including new methodologies or data. This step brings the reproduction exercise a step closer to *replication*.

Figure 2: Steps for reproduction



Suggested level of effort:

- Graduate research:	5%	10%	5%	10%	70%
- Graduate course:	10%	25%	20%	40%	5%
- Undergrad thesis:	10%	30%	40%	20%	0%

Figure 2 depicts suggested levels of effort for each stage of the exercise depending on the context in which you are performing a reproduction. This process need 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 a paper's level of reproducibility.

Recording the results of the exercise

You will be asked to record the results of your reproduction as you progress through each stage.

In *Stage 1: Scoping*, complete **Survey 1**, where you will declare your paper of choice and the specific display item(s) and specifications on which you will focus for the remainder of the exercise. This step may also involve writing a brief 1-2 page summary of the paper (depending on your instructor or goals).

In *Stage 2: Assessment*, you will inspect the paper's reproduction package (raw data, analysis data, and code), connect the display item to be reproduced with its inputs, and assign a reproducibility score to each output.

In *Stage 3: Improvement*, you will try to improve the reproducibility of the selected outputs by adding missing files, documentation, and report any potential changes in the level of reproducibility. Use **Survey 2** to record your work at Stages 2 and 3 (you will receive access instructions for Survey 2 when you submit Survey 1).

In *Stage 4: Robustness Checks*, you will assess different analytical choices and test possible variations. Use **Survey 3** to record your work at this stage.

Reproduction Strategies

Generally, a reproduction will begin with a thorough reading of the study being reproduced. However, subsequent steps may follow from a *reproduction strategy*. For example, a reproduction may closely follow the order of the steps outlined above. This might entail the reproducer first choosing a set of results whose pproduction they are interested in assessing or understanding, completely reproducing these results to the extent possible, and then making modifications to the reproduction package. Another potential strategy could be for the reproducer to develop potential robustness checks or extensions while reading the

study, which would lead to the definition of a set of results to be assessed via reproduction. Yet another reproduction strategy may be for the reproducer to seek out a paper that uses a particular dataset to which they have access or an interest in using, reproducing the results that use that dataset as an input, then probing the robustness of the results to various data cleaning decisions.

The various uses of reproduction makes the number of potential reproduction strategies quite large. In choosing or designing a reproduction strategy, it is helpful to clearly identify the goal of the reproduction. In all of the examples laid out in the paragraph above, the order in which the steps of the reproduction exercise are taken is at least partially determined by what the reproducer hopes to get from the exercise. The structure provided in these guidelines, together with a clear reproduction goal, can facilitate the implementation of an efficient reproduction strategy.

Chapter 1

Scoping

In this stage, you will define the scope of your exercise by declaring a paper and the specific output(s) on which you will focus. 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**.

It is likely that you will choose a declared paper based on whether or not you can locate its 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 6) 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 ACRE database** (currently under development).

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. If materials are available, you will read the paper and declare the scope of the reproduction exercise. You can expect to spend between 1-3 days in this Scoping stage, though this may vary based on the length and the complexity of the paper, and the availability of reproduction materials.

Use Survey 1 to record your work in this stage.

1.1 From candidate to declared paper

At this point of the exercise, you are *only validating 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 mean that you have found your declared paper).

1. Check whether previous reproduction attempts have beeen recorded in the ACRE Database for the paper (more on the ACRE Database in the next section).
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 5.
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 exericse 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 #) 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 ACRE database* 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 ACRE Database

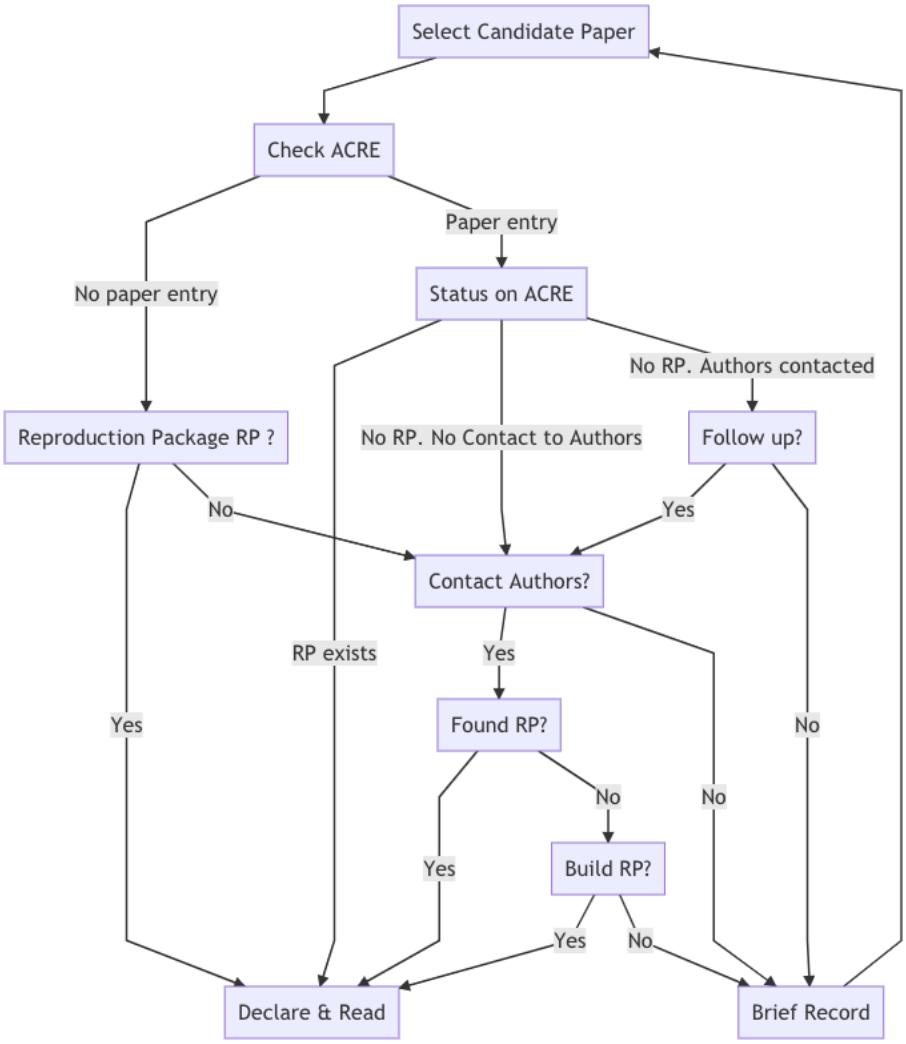
If the ACRE database contains previous reproduction attempts of the paper, you will see a report card with the following information:

> **Box 1:** Summary Report Card for ACRE Paper Entry
> **Title:** Sample Title

Authors: Jane Doe & John Doe
Original Reproduction Package Available: URL/No [What does this mean? Add some context]. [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 ACRE Reproductions: Yes/No/Unknown
Open for reproductions: Yes/No [Same as above: what does this mean? Add more context].

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 ACRE database following the example above.

View Decision Tree To Select Paper (Emma: add title and solve bug with svg)



1.2 Scoping your declared paper

Once you have identified your declared paper, get familiarized with it and choose the specific output(s) on which you will focus for the remainder of the exercise.

1.2.1 Read and summarize the paper

Depending on how much time you have, we recommend that you write a short (1-2 page) summary of the paper. This will help remind you of the key elements to focus on for the reproduction, and demonstrate your understanding of the paper (for yourself and others like your instructor or advisor).

When reading or summarizing the paper, try to answer the following questions:

- Would you classify the paper's scientific claims as mainly focused on estimating a causal relationship, estimating/predicting a descriptive statistic of a population, or something else?
- How many scientific claims (descriptive or causal) are investigated in the paper?
- What is the population for which the estimates apply?
- What is the population that is the focus of the paper as a whole?
- What are the main data sources used in the paper?
- How many display items are there in the paper (tables, figures, and inline results)?

- What is the main statistical or econometric method used to examine each claim?
- What is the author's preferred specification (or yours, if the authors are not clear)?
- What are some robustness checks for the preferred specification?

1.2.2 Record scope of the exercise

By now you should have a fairly good understanding of the paper's content. You do not, however, need to have spent any time reviewing the reproduction package in detail.

At this point, you should clearly specify which part of the paper 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 **??**. If you plan to scope more than one claim, *we strongly recommend starting with just one* and recording your results. You can then initiate another record in ACRE later for the second (or third, fourth, etc.) claim to reproduce using the materials and knowledge you developed in the first exercise. You can, however, reproduce more than one claim if you are already familiar with the paper.

In the Assessment stage, the reproduction will be centered around the display item(s) that contain the specification you indicate at this point.

Declare specific main estimates to reproduce.

Identify a scientific claim and its corresponding preferred specification, and record its magnitude, standard error, and location in the paper (page, table #, and table row and column). If the authors did not explicitly chose a particular estimate, you will be asked to select one. In addition to the preferred estimate, reproduce up to five estimates that correspond to alternative specifications of the preferred estimate.

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.

These are the elements you will need for the Scoping stage. **You now have all the elements necessary to complete Survey 1.**

1.3 Identify your relevant timeline.

Before you begin working on the three main stages of the reproduction exercise (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 longer class project that may take a month to complete or 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, and acknowledgin that less experienceed 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.

Emma: please write this table using R and KableExtra

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)

5% (1 day)

5% (5 days)

Assessment

35%

25%

15%

Improvement

25%

0%

40%

20%

30%

Robustness

25%

5%

25%

25%

```
library(tidyverse)
library(knitr)
library(kableExtra)
temp_eval <- TRUE
options(tinytex.verbose = TRUE)
```

Chapter 2

Assessment

In this stage, you will review and describe in detail the available reproduction materials, and assess levels of computational reproducibility for the selected outputs, as well as for the overall paper. This stage is designed to record as much of the learning process behind a reproduction as possible to facilitate incremental improvements, and allow future reproducers to pick up easily where others have left off.

First, you will provide a detailed description of the reproduction package. 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.

In the *Scoping* stage, you declared a paper, identified the specific claims you will reproduce, and recorded the main estimates that support the claims. In this stage, you will identify all outputs that contain those estimates. You will also decide if you are interested in assessing the reproducibility of that entire output (e.g., "Table 1"), or will assess only a pre-specified estimates (e.g., "rows 3 and 4 of Table 1"). Additionally, you can include other outputs of interest.

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.

2.1 Describe the inputs.

This section explains how to list *all* input materials found or referred to in the reproduction package. First, you will identify data sources and connect them with their raw data files (when available). Second, you will locate and provide a