# [MC number] [Manuscript Title] Validation and Replication results

INSTRUCTIONS: This report can be used by Third Party Replicators. It is NOT a complete report, as the AEA staff will add additional information regarding data sources.

INSTRUCTIONS: For more information on what it means to be a Third Party Replicator, please see the AEA policy at <https://www.aeaweb.org/journals/data/policy-third-party>.

INSTRUCTIONS: Once you’ve read these instructions, DELETE THESE AND SIMILAR LINES. Also delete lines that include “{{ SOME-TEXT }}”.

INSTRUCTIONS: In the above title, replace [Manuscript Title] with the actual title of the paper, and [MC number] with the Manuscript Central number (e.g., AEJPol-2017-0097)

INSTRUCTIONS: Go through the steps to download and attempt a replication. Document your steps here, the errors generated, and the steps you took to alleviate those errors. This includes apparently minor steps, such as adjusting directories, or installing packages, any deviations from the README. If the README did not specify a certain step or action, describe why you chose that action, and whether it should be detailed in the README. All figures and tables should be evident once you are done, i.e., saved to disk. If this is not done in code, please add code to do so.

INSTRUCTIONS: You should NOT communicate with the authors of the manuscript on anything other than how to obtain access to the data. If some communication is absolutely necessary, do so by email, and copy the Data Editor at dataeditor@aeapubs.org. Also note such communication in this report.

INSTRUCTIONS: To compare images and tables, annotated screenshots from the manuscript can be helpful for comparison, highlighting where differences were found. However, comply with all disclosure avoidance/ results release rules.

INSTRUCTIONS: Once completed, send this file, any allowed log files and output, back to the requesting editor.

INSTRUCTIONS: Fill out the following section:

## Affirmations

* ☐ I have identified any deviations from the PUBLIC README, noted any bugs (even minor) that needed to be fixed.
* ☐ I have not had any communications with the authors of this manuscript on this package ever, other than to coordinate access to data; OR
* ☐ I have had communications with the authors of this manuscript, and I copied the AEA Data Editor on all such communication; OR
* ☐ I have had communications with the authors of this manuscript, without involving the AEA Data Editor.

## Data description

### Analysis Data Files

INSTRUCTIONS: Separately, identify any analysis data file provided. Analysis data files are produced by code in the deposit from data sources. Not every deposit will have these. Analysis data files do not need to be cited, because they are a key part of the replication package.

* ☐ No analysis data file mentioned
* ☐ Analysis data files mentioned, not provided (explain reasons below)
* ☐ Analysis data files mentioned, provided. File names listed below.
* ☐ Analysis data files not mentioned, but provided.

Example:

./Output\_Empirical/data/regression\_main.dta

## Stated Requirements

INSTRUCTIONS: The authors may have specified specific requirements in terms of software, computer hardware, etc. Please list them here. This is **different** from the Computing Environment of the Replicator. You have the option to amend these with unstated requirements later. If all requirements are listed, check the box “Requirements are complete”.

* ☐ No requirements specified
* ☐ Operating system used:
  + ☐ Windows 10/11
  + ☐ MacOS
  + ☐ Linux
  + ☐ Windows Server
  + ☐ Not specified
* ☐ Software Requirements specified as follows:
  + Software 1
  + Software 2
* ☐ Computational Requirements specified as follows:
  + Cluster size, etc.
* ☐ Time Requirements specified as follows:
  + Length of necessary computation (hours, weeks, etc.)
* ☐ Requirements are complete.

INSTRUCTIONS: If the requirements are NOT complete, please leave this line in. If UNSURE, leave this line in:

For missing requirements, see the list of required changes in the [**FINDINGS**](#findings) section.

INSTRUCTIONS: If easier, simply copy-and-paste the authors’ stated requirements here:

## Code description

INSTRUCTIONS: Review the code (but do not run it yet). Identify programs that create “analysis files” (“data preparation code”). Identify programs that create tables and figures. Not every deposit will have separate programs for this.

INSTRUCTIONS: Identify all **Figure, Table, and any in-text numbers**. Create a list, mapping each of them to a particular program and line number within the program (use [this template](code-check.xlsx)). Commit that list. You will come back to the list in your findings. IN THIS SECTION, point out only a summary description, including of shortcomings. E.g.

INSTRUCTIONS: For example, you could write “There are four provided Stata do files, three Matlab .m files, including a”master.do“.” INSTRUCTIONS: And you could list the issues you encounter: INSTRUCTIONS: - Table 5: could not identify code that produces Table 5 INSTRUCTIONS: - Neither the program codes, nor the README, identify which tables are produced by what program.

{{ programs-summary.txt }}

The full list of programs provided can be found in the Appendix.

INSTRUCTIONS: Verify the appendix, the repository. If the list is missing, generate it by hand (instructions in the appendix).

* ☐ The replication package contains a “main” or “master” file(s) which calls all other auxiliary programs.

INSTRUCTIONS: If the above checkbox for “main” file is NOT checked, leave the following SUGGESTION in the report!

[SUGGESTED] We strongly advise the use of a single (or a small number of) main control file(s) to automatically reproduce all figures and tables in the paper, without manual interaction.

NOTE: In-text numbers that reference numbers in tables do not need to be listed. Only in-text numbers that correspond to no table or figure need to be listed.

## Computing Environment of the Replicator

INSTRUCTIONS: You are now running code in *PartB*. Please move the *PartB* sub-task to *Code is running*!

INSTRUCTIONS: If there is an external reproducibility report (e.g., World Bank, cascad), please include it in the repository, and mention it here, and delete the other lines.

* ☐ The reproducibility check was conducted by a trusted third-party, see their appended report for details.

INSTRUCTIONS: This might be automated, for now, please fill in manually. Remove examples that are not relevant, adjust examples to fit special circumstances. Some of this is available from the standard log output in Stata or R. Some frequently used details are below. Some of these details can be found as follows:

* (Windows) by right-clicking on “This PC”
* (Mac) Apple-menu > “About this Mac”
* (Linux) see code in tools/linux-system-info.sh

Some options are listed below. Choose the one that applies.

* Mac Laptop M2 MacBook Pro, MacOS 10.14.6, 8 GB of memory, (M2 Processor, 8 cores)
* RedCloud Servers: Intel Xeon Processor (Skylake) 3.09 GHz (18 processors), Installed Ram 176 GB, 64-bit operating system, x64-based processor
* CCSS Cloud: Windows Server AMD EPYC 7763 64-Core Processor 2.44 GHz, 128GB
* BioHPC Linux server, Rocky Linux 9.0, AMD Opteron Processor 6380 / 16 cores/ 125 GB memory (adjust as necessary from output of linux-system-info.sh)
* WholeTale (describe the environment chosen)
* CodeOcean (describe the type of capsule chosen) Intel(R) Xeon(R) Platinum 8259CL CPU @ 2.50GHz (16 cores), 128 GB Memory
* Bitbucket Pipelines, “Ubuntu 20.04.2 LTS”, Intel(R) Xeon(R) Platinum 8259CL CPU @ 2.50GHz, 8 cores, 9GB memory
* Codespaces, “Ubuntu 20.04.2 LTS”, Intel(R) Xeon(R) Platinum 8272CL CPU @ 2.60GHz, 2 core, 8GB/ 4-core, 16GB/ 8-core, 32 GB/ 16-core, 64GB (choose appropriately)

## Replication steps

INSTRUCTIONS: If there is an external reproducibility report (e.g., World Bank, cascad), please include it in the repository, and mention it here, and delete the other lines.

* ☐ The reproducibility check was conducted by a trusted third-party, see their appended report for details.

INSTRUCTIONS: provide details about your process of accessing the code and data.

* Do NOT detail things like “I save them on my Desktop”.
* DO describe actions that you did as per instructions (“I added a config.do”)
* DO describe any other actions you needed to do (“I had to make changes in multiple programs”), without going into TOO much detail. (Link to the log file in the JIRA comments of this case!)

BUT:

* DO provide ENOUGH detail so that an author, without access to the logs, can understand what needed to be fixed, including a copy-paste of the error message.
* DO commit to git before EACH new run with corrected code.
* DO (after all debugging is completed) a full run through the data, top-to-bottom, once all bugs are fixed, using the approriate method (command line or right-click).

Example:

* Downloaded code and data from openICPSR provided.
* Added the config.do generating system information.
* Ran code as per README, but the third step did not work. (describe the problem HERE, but link to the logfile in the JIRA comments) (commit)

error: command distinct unknown

* Had to add undocumented package distinct to the install portion of the config file. Ran again.
* Code failed because of a typo in the name of the file “superdata.dta” (was: superdta.dta). Fixed. Ran again.
* Code ran to completion.

## Findings

INSTRUCTIONS: Describe your findings both positive and negative in some detail, for each **Data Preparation Code, Figure, Table, and any in-text numbers**. You can re-use the Excel file created under *Code Description*. When errors happen, be as precise as possible. For differences in figures, provide both a screenshot of what the manuscript contains, as well as the figure produced by the code you ran. For differences in numbers, provide both the number as reported in the manuscript, as well as the number replicated. If too many numbers, contact your supervisor.

INSTRUCTIONS: Even when there is an external reproducibility report, summarize the findings here.

### Missing Requirements

INSTRUCTIONS: If the replication package contains Stata programs run tools/Stata\_scan\_code/scan\_packages.do, ensuring that you update the global codedir first. If the data is accessible, add any packages not mentioned in the README to the config.do and paste the excel output as a table below. If the data is restricted-access and not obtainable in a reasonable amount of time, paste the excel output as a table below.

INSTRUCTIONS: If it turns out that some requirements were not stated/ are incomplete (software, packages, operating system), please list the *missing* list of requirements here. Remove lines that are not necessary. If the stated requirements are complete, delete this entire section, including the [REQUIRED] tag at the end.

* ☐ Software Requirements
  + ☐ Stata
    - ☐ Version
    - Packages go here
  + ☐ Matlab
    - ☐ Version
  + ☐ R
    - ☐ Version
    - R packages go here
  + ☐ Python
    - ☐ Version
    - Python package go here
  + ☐ REPLACE ME WITH OTHER
* ☐ Computational Requirements specified as follows:
  + Cluster size, disk size, memory size, etc.
* ☐ Time Requirements
  + Length of necessary computation (hours, weeks, etc.)

[REQUIRED] Please amend README to contain complete requirements.

You can copy the section above, amended if necessary.

### Data Preparation Code

Examples:

* Program 1-create-data.do ran without error, output expected data
* Program 2-create-appendix-data.do failed to produce any output.

### Tables and Figures

INSTRUCTIONS: Insert the filled-out code-check.xlsx here (complete the column Reproduced?), using the VS Code Plugins [Excel to Markdown table](https://marketplace.visualstudio.com/items?itemName=csholmq.excel-to-markdown-table). Then describe in more detail the issues that may have arisen.

Examples:

* Table 1: Looks the same
* Table 2: (contains no data)
* Table 3: Minor differences in row 5, column 3, 0.003 instead of 0.3

INSTRUCTIONS: For tables, simple comparisons can be listed out as above. More complex differences can be described by using screenshots of the original table and the reproduced table, highlighting the differences.

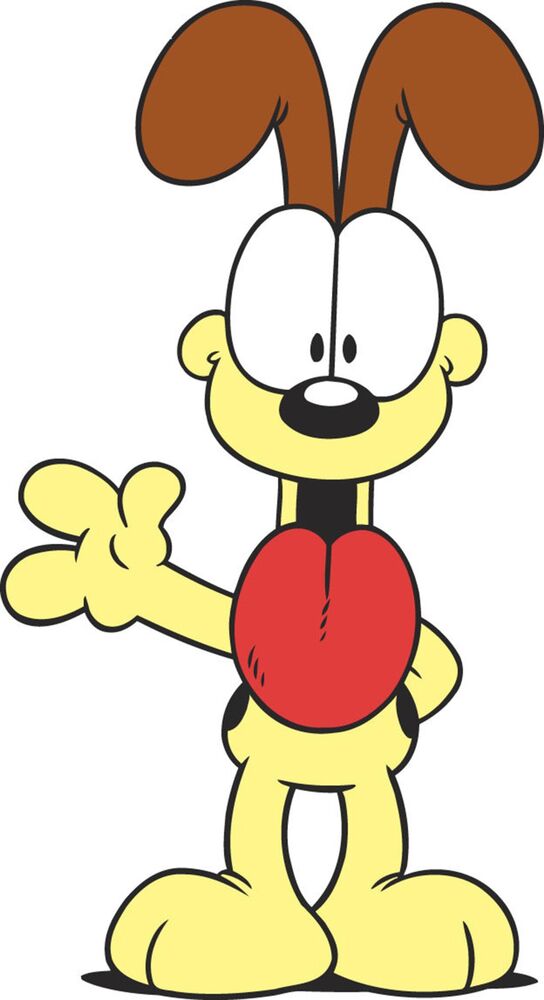
INSTRUCTIONS: Please provide a comparison with the paper when describing that figures look different. Use a screenshot for the paper, and the graph generated by the programs for the comparison. Reference the graph generated by the programs as a local file within the repository.

Example:

* Figure 1: Looks the same
* Figure 2: no program provided
* Figure 3: Paper version looks different from the one generated by programs:

Paper version: 

Figure 3 generated by programs:



Replicated version

### In-Text Numbers

INSTRUCTIONS: list page and line number of in-text numbers. If ambiguous, cite the surrounding text, i.e., “the rate fell to 52% of all jobs: verified”.

[ ] In-text numbers not verified.