



Safety First: Perceived Risk of Street Harassment and Educational Choices of Women

Third Submission: RR_IND_2024_244

Ceren Huryol, Ankriti Singh, Maria Reyes Retana Torre

reproducibility@worldbank.org

June 12, 2026

This review verifies the reproducibility of the exhibits included in the paper “*Safety First: Perceived Risk of Street Harassment and Educational Choices of Women*”.

Contents in this review:

1. Main findings
2. List of exhibits and reproducibility status
3. Reproduction Environment

Main findings

- **Every exhibit in the main section of the paper and 10 randomly selected exhibits in the appendix were reproduced accurately**
- The code was partially reproduced on a new computer due to the extensive runtime of one Matlab script (`mc_simulations_test.m`, approximately 2 weeks), which was not executed by the replicators. See below for details.
- The following steps were executed on a new computer:
 1. Updating the working directory in line 26 of `master_1.do` and running it.
 2. Running `master_2.m` in Matlab, except for `mc_simulations_test.m`.
 3. Updating the working directory in line 26 of `master_3.do` and running it.
- The code takes approximately 20 days to run.
- We conducted our reproducibility analysis based on the paper shared by the authors.
- **Verification Process and Data Handling:**
 - The reproducibility package relies on 4 types of data: open, accessible (public but not included in the reproducibility package), limited-access, and restricted data (available only from the data owners).
 - Due to the extensive runtime of `mc_simulations_test.m` (approximately 2 weeks), this script was not executed during reproducibility verification. The replicators ran Part 1 (`master_1.do`) and the Matlab component except for `mc_simulations_test.m`, shared the required inputs with the authors, who ran `mc_simulations_test.m` and returned the outputs. The replicators used these outputs to run Part 3 (`master_3.do`) and verify the final exhibits.

- Open data is not included in the public reproducibility package due to file size constraints.
- Reviewers used forthcoming/accessible/restricted data provided directly by the authors to conduct the reproducibility verification, and this is not included in the public reproducibility package.
- *data_hash_report.csv* lists the SHA256 hashes of all files in the Data folder to support data integrity checks. Users who acquire the forthcoming/accessible/limited-access/restricted data can use this file to verify that the data has not been altered.
- **Reproducibility Summary:**
 - **Data:** Some data is restricted and has not been included in the reproducibility package. For more details, please refer to the README file.
 - **Code:** All code files (from cleaning to analysis) are included in the reproducibility package.
 - **Outputs:** Some outputs are not generated by code (created manually in ArcGIS) but instructions for producing these outputs are included in the README file.
 - **Dependencies environment:** The reviewers created a new environment using the latest versions of dependencies available at the moment of the review.

List of exhibits and reproducibility status

Results in the Main Section of the Paper

- **Figure 1a** Reproduced.
- **Figure 1b** Reproduced.
- **Figure 2** Does not apply. Figure does not represent analysis results.
- **Figure 3a** Reproduced.
- **Figure 3b** Reproduced.
- **Figure 4** Reproduced. Figures were manually prepared in ArcGIS.
- **Figure 5** Reproduced. Figures were manually prepared in ArcGIS.
- **Figure 6** Reproduced. Figure was manually prepared in ArcGIS.
- **Figure 7a** Reproduced.
- **Figure 7b** Reproduced.
- **Figure 7c** Reproduced.
- **Figure 7d** Reproduced.
- **Figure 8a** Reproduced.
- **Figure 8b** Reproduced.
- **Figure 8c** Reproduced.

- **Figure 8d** Reproduced.
- **Figure 8e** Reproduced.
- **Figure 9** Reproduced.
- **Table 1** Reproduced.
- **Table 2** Reproduced.
- **Table 3** Reproduced.
- **Table 4** Reproduced.
- **Table 5** Reproduced.
- **Table 6** Reproduced.

Results in the Annex

For the Appendix, we did not review every exhibit. Instead, we randomly selected 10 exhibits from the remaining datasets to assess the appendix. Our review was based on those 10 exhibits. Since they were chosen randomly, we are operating under the assumption that if all randomly selected exhibits are reproducible, then the rest should be as well. The seed used to generate the random selection was 813433, the Stata version used was 18.5, and the exhibits selected are listed below:

- **Figure A3** Reproduced.
- **Figure A6** Reproduced.
- **Figure A8** Reproduced.
- **Figure A10** Reproduced.
- **Figure A12** Reproduced.
- **Figure A13** Reproduced.
- **Table B2** Reproduced.
- **Table B6** Reproduced.
- **Table B10** Reproduced.
- **Table B13** Reproduced.

Reproduction Environment

Paper exhibits were reproduced on a computer with the following specifications:

- OS: Windows 11 Enterprise
- Processor: INTEL(R) XEON(R) PLATINUM 8562Y+ 2.80 GHz (2 processors)
- Memory available: 32.0 GB
- Software version: Stata 18.0 MP; ArcGIS Pro 3.5.3; Matlab 2024a

Portion of the Matlab component was reproduced on a computer with the following specifications:

- OS: Windows 10 Enterprise
- Processor: Intel Xeon Gold 6132 CPU @ 2.60GHz (16 cores, 16 logical processors)
- Memory available: 128.0 GB
- Software version: Matlab 2024b